AFRL-RH-WP-TR-2012-0078

EVALUATION OF DOMAIN-SPECIFIC COLLABORATION INTERFACES FOR TEAM COMMAND AND CONTROL TASKS

**Gregory J. Funke, Allen W. Dukes, James M. Hyson,
Brent T. Miller, Sheldon M. Russell, Benjamin A. Knott, and Scott Galster**

Air Force Research Laboratory

May 2012

Distribution A: Approved for public release; distribution unlimited.

**AIR FORCE RESEARCH LABORATORY
711 HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2012-0078 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FUNKE.GREGORY.J.1283346494
Digitally signed by FUNKE.GREGORY.J.1283346494
DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USAF, cn=FUNKE.GREGORY.J.1283346494
Date: 2012.06.04 16:26:14 -04'00'

GREGORY J. FUNKE
Work Unit Manager
Collaborative Interface Branch
Warfighter Interface Division

RUSSELL.WILLIAM.E.JR.1007068057
Digitally signed by RUSSELL.WILLIAM.E.JR.1007068057
DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USAF, cn=RUSSELL.WILLIAM.E.JR.1007068057
Date: 2012.06.04 16:30:43 -04'00'

WILLIAM E. RUSSELL
Chief, Collaborative Interface Branch
Warfighter Interface Division

STROPKI.MICHAEL.A.1230098987
Digitally signed by STROPKI.MICHAEL.A.1230098987
DN: c=US, o=U.S. Government, ou=DoD, ou=PKI, ou=USAF, cn=STROPKI.MICHAEL.A.1230098987
Date: 2012.06.05 12:25:00 -04'00'

MICHAEL STROPKI
Chief, Warfighter Interface Division
Human Effectiveness Directorate
711Human Performance Wing

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*
01/05/12

2. REPORT TYPE
INTERIM

3. DATES COVERED *(From - To)*
24 October 2008 – 3 January 2012

4. TITLE AND SUBTITLE
EVALUATION OF DOMAIN-SPECIFIC COLLABORATION INTERFACES FOR TEAM COMMAND AND CONTROL TASKS

5a. CONTRACT NUMBER
In House

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Gregory J. Funke, Allen W. Dukes, James M. Hyson, Brent T. Miller, Sheldon M. Russell, Benjamin A. Knott, and Scott M. Galster

5d. PROJECT NUMBER
7184

5e. TASK NUMBER
25

5f. WORK UNIT NUMBER
71840877

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Applied Neuroscience Branch
Wright-Patterson Air Force Base, OH 45433

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711 Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Applied Neuroscience Branch
Wright-Patterson Air Force Base, OH 45433

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
711 HPW/RHCP

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RH-WP-TR-2012-0078

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A: Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
88 ABW Cleared 06/13/2012; 88ABW-2012-3343. Report contains color.

14. ABSTRACT
Two experiments were conducted to evaluate the impact of a domain-specific digital whiteboard tool and a tactical resource management display on team performance efficiency in a simulated command and control (C2) air defense scenario. The two experiments differed primarily in the participant population employed – in the first experiment novices (college students) with no C2 experience served as participants, and in the second experienced domain experts were the participants. During the experiment participants assumed the roles of weapons directors responsible for coordinating offensive counter-air, defensive counter-air, and air refueling with their teammates. The results of both experiments suggest that the collaboration technologies examined are likely to reduce reliance on radio communication without adversely affecting team performance, and under some circumstances they may provide small performance gains.

15. SUBJECT TERMS
Command and control (C2), team research, collaboration technology, digital whiteboard

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
Unclassified	Unclassified	Unclassified	SAR	46	Gregory J. Funke **19b. TELEPHONE NUMBER** *(Include Area Code)* (937) 938-3601

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK.

TABLE OF CONTENTS

Section **Page**

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

1.0 INTRODUCTION

As U.S. military operations continue in Iraq and Afghanistan and new commitments arise in the broader Middle East, it becomes increasingly important to provide a common operational picture to joint forces within and across theaters of operations. Military operations require a high degree of communication, collaboration, and coordination (of personnel, assets, etc.) to achieve mission success. Proponents of network-centric operations (e.g., Alberts & Hayes, 2003; Hayes, 2004) have proposed that integration and performance of distributed teams may be facilitated through emerging collaboration technologies, such as email, instant messaging ("chat"), virtual whiteboards, and video and desktop conferencing applications (Boiney, 2005). Those authors argue that such technologies could engender a degree of command decentralization resulting in increased situational awareness and task flexibility (Alberts & Hayes, 2003). While the past two decades have seen rapid advances in collaboration technology development, researchers are still exploring the impact of these tools on team performance, communication, and workload in military environments (see e.g., Baltes et al., 2002, and Hertel et al., 2005, for reviews).

A significant dimension of combat operations is command and control (C2), which has been previously defined as "the exercise of authority and direction by a properly designated commander over assigned and attached forces in the accomplishment of the mission" (U.S. Department of Defense, 2001). Air battle management (ABM) constitutes the command and control of air-to-air and air-to-ground operations, and involves the direction and implementation at the tactical level of operational air tasking orders (Vidulich, Bolia, & Nelson, 2004). Examples of ABM operations include the control of assets engaged in offensive and defensive operations, air refueling, and air mobility missions. To achieve the objectives of these operations, military personnel must work within small distributed teams to receive and transmit information across various platforms, make tactical decisions, coordinate actions, and disseminate plans (Knott, Bolia, Nelson, & Galster, 2006).

Generally, oral radio communication has been the dominant collaboration technology employed by teams in ABM operations (Vidulich et al., 2004). While effective, radio communication is subject to a number of limitations. First, communication is subject to serial transmission (e.g., only one person can talk at a time) and limited transmission bandwidth. In addition, the low quality of radio transmissions and the presence static interference is likely to reduce speech intelligibility (Bolia, Nelson, Vidulich, Simpson, & Brungart, 2005). Second, radio messages are transient and do not include an archive of past communications, which prevents operators from "looking back" to confirm or refresh their understanding of dialogue, resulting in missed information, misinformation, and repeated requests for information. In addition, it is possible that reliance on any single communication modality may stifle a team's flexibility and responsiveness to dynamic changes in operational environments (e.g., if radio communications are disrupted or compromised; Alberts & Hayes, 2003).

Based on these concerns, two studies were conceived to examine the effects of collaboration technologies designed to supplement radio communication on team performance, radio traffic, and perceived workload in simulated air battle management. The first experiment utilized trained novices in a controlled laboratory simulation; the second was similar to the first, but participants were ABM domain experts.

1.1. Supplemental Collaboration Technologies

1.1.1. Virtual Whiteboard

Cognitive theories relating the utilization, storage, and retrieval of verbal and spatial information, such as Wickens' (1980) multiple resource theory or Baddeley's (1986) model of working memory, propose separate encoding and processing of each. An important implication of these models is that verbal communications concerning spatial information requires both the sender and receiver of a message to transform one representational form to the other (i.e., spatial to verbal for the speaker, and verbal to spatial for the listener). However, as noted by Wickens, Vidulich, and Sandry-Garza (1984), communication of spatial information is often delivered and received more effectively through a visual, rather than verbal, medium. Consequently, a collaboration technology such as a virtual whiteboard, which affords teams the ability to represent and transmit spatial information pictorially, may positively impact team performance.

Previous research concerning the utility of a virtual whiteboard for C2 tasks generally supports this idea, in that teams report preferring to communicate spatial information using a whiteboard, and that access to a whiteboard can improve team performance (e.g., Schwartz, Knott, & Galster 2008; Vincent et al., 2009). However, such benefits likely only emerge when monitoring the whiteboard does not interfere with ongoing task duties, and when participants are given sufficient practice with the tool (Funke & Galster, 2009).

1.1.2. Team Resource Display

A key tenet of network-centric operations is that mission-relevant information should be accessible to decision-makers at all levels of an organization (Alberts & Hayes, 2003; Hayes, 2004). Current and pending technological improvements to C2 systems aim to increase the surveillance, control, and communication capabilities of C2 platforms by enabling greater information sharing between military assets than is currently feasible (e.g., Jeziorski, 2008; Sloan, 2009). As mentioned previously, the suggested advantage of this approach is that information availability will facilitate situation awareness and improve the adaptability of team operations by enabling teams with shared information and empowering them through command decentralization. Shared awareness among team members may foster adaptability to changing situations by generating common interpretations of evolving environmental constraints (Salas, Cooke, & Rosen, 2008).

A potential secondary benefit of the net-centric approach may be a reduced need for operator communication. For example, in an ABM context, relatively simple, domain-specific resource displays, designed to convey information about team fuel and weapons status, may increase team situation awareness, but also reduce team communication regarding those resources. However, it is also possible that such tools may increase operator workload if the information conveyed by the display is not easily accessible, or if it creates an additional monitoring burden (Parasuraman & Riley, 1997).

2.0 EXPERIMENT 1

2.1. Introduction

The influence of collaboration tools on ABM operations is still being explored within the research literature, even as these technologies are increasingly deployed in combat operations (e.g., Hayes, 2004). The purpose of Experiment 1 was to evaluate the impact of a domain-specific digital whiteboard tool and a team resource display on team performance in a laboratory setting using novice operators in a simulated ABM environment. This step was important for investigating the utility of the collaboration tools under consideration in conditions of more rigorous control than would have been possible given the constraints of an operational environment. However, following the suggestions of Cooke and Shope (2004), care was taken to ensure domain applicability and generalization by including domain-relevant characteristics, such as high mental and temporal demands and moderate team interdependence, in the simulation.

Two research questions were explored in Experiment 1. First, how does team communication change with access to supplemental collaboration tools, and second does that change impact team performance and workload? As discussed previously, one possible outcome is that additional communication outlets will allow teams to exchange information more efficiently and effectively, potentially resulting in improved team performance and reduced operator workload. Conversely, increasing the number of communication channels may cause operators to divide attention between performing the primary ABM task and monitoring for information updates, decrementing team performance and increasing workload. In pursuit of these experimental questions, several indices of team communication, performance, and operator workload were recorded and assessed in Experiment 1.

Based on the reviewed literature, it was expected that the dual availability of a virtual whiteboard and resource display would improve team performance on the primary ABM task, and on a secondary auditory monitoring task. It was also hypothesized that collaboration tool availability would decrease the overall number and duration of radio transmissions produced by teams, though no specific hypothesis was made regarding the effect of the tools on the semantic content of team communications. Finally, it was hypothesized that tool availability would reduce ratings of operator workload. Together, these findings would support the further development of supplemental collaboration technologies for use in ABM operations.

2.2. Methods

2.2.1. Participants

Sixteen men between the ages of 18 and 28 years old (*M* = 21.86, *SD* = 2.85) served as participants in this experiment. Participants were students recruited from local universities and were compensated for their participation. The experiment also included six confederates. Confederates were compensated at the same rate as participants. In total, the experimental sample included eight teams; each team consisted of two participants and three confederates.

2.2.2. Experimental Design

A within-subjects design was employed, with two team communication conditions (standard, augmented) combined factorially with two resource display conditions (present, absent) yielding four experimental conditions. Each experimental team completed two mission trials in each condition, for a total of eight experimental trials. Team communication and resource display were both blocked factors, such that team communication condition was organized as two-trial blocks, within the larger four-trial blocks of the resource display condition. The order of presentation of trial conditions was counterbalanced across teams.

Dependent measures included in this experiment comprised indices of team performance in a simulated air defense task; performance on an auditory monitoring task; frequency, duration, and content of team communications; and several measures of subjective workload.

2.2.3. Apparatus

2.2.4. Workstations

This experiment required five workstation computers for the participants and confederates, three “observer” computers for the experimenters, one Synthetic Task Environment (STE) server, one computer hosting a Structured Query Language (SQL) Server database, and one domain controller. The workstation computers were equipped with a single Dell 1703FPs 17 inch LCD monitor, a Logitech *QuickCam for Notebooks Pro* web camera (model 960-000045), a standard optical mouse, and a standard keyboard. The five workstation computers and the primary observer station each had a virtual machine configured to provide the use of the Linux based STE client. This implementation of a virtual machine was necessary to enable participants to interact with the Linux-based STE and the Windows environment. The primary observer station also hosted software which allowed experimenters to implement the conditions of each trial. The remaining two observer stations hosted additional data recording applications (detailed below). All computers were networked using a Netgear GS748T gigabit switch which provided standard TCP/IP Ethernet connectivity. A complete list of the hardware specifications for each computer employed in this experiment is displayed in Table 1.

Table 1. Hardware specifications for the eleven computers employed in Experiment 1.

Computer	Quantity	Manufacturer	Model	Processer	Operating System	RAM	Network
Participant workstations	5	Dell	Optiplex GX270	Intel Pentium 4 2.8 GHz	MS XP Professional	2 GB	1 Gbps
STE server	1	Dell	Precision 340	Intel Pentium 4 2.0 GHz	Red Hat Linux 9.0	1 GB	.1 Gbps
Primary observer station	1	HP	Compaq dc7100	Intel Pentium 4 3.2 GHz	MS XP Professional	1 GB	1 Gbps
Secondary observer station	2	Visionman	V133-2335	Intel Celeron 2.4 GHz	MS XP Professional	.5 GB	.1 Gbps
SQL server	1	5 O'Clock	Custom	Intel Xeon 3.06 GHz	MS Windows Server 2003	2 GB	1 Gbps
Web service and domain controller	1	Dell	PowerEdge 400SC	Intel Pentium 4 3.2 GHz	MS Windows Server 2003	1 GB	1 Gbps

Note. HP = Hewlett-Packard, MS = Microsoft.

During the experiment, teammates communicated with each other using simulated radio headsets. Each workstation was equipped with a set of Sennheiser HD250 Linear II headphones and a Sennheiser HMD 224 microphone. Prior to starting the experiment, the microphone at each workstation was calibrated to each speaker (participant or confederate) using WaveSurfer (version 1.8.5; Sjölander & Beskow, 2005), an audio editing application. An Applied Research Technology (ART) HPFX Headphone Monitor System was used to transmit team communications from the microphone into the Windows environment. General Dynamics C4 Systems, Inc.'s *ModIOS Voice Communicator* application (version 2.3.4, 2002) then converted the speech information into Distributed Interactive Simulation Protocol Data Units (DIS PDUs) and the information was broadcasted over the network to teammates. Upon receiving a teammate's communication, ModIOS translated the DIS PDUs back to speech, which was relayed to participants through their headphones. In conjunction with ModIOS, the Warfighter Communication Assessment System (WCAS; 2005), developed by the Air Force Research Laboratory (AFRL), was used to capture DIS PDUs transmitted across the network, convert them to .wav files, and save them to the computer's hard drive. In addition, the program *DISlog* (part of the DIScretion software suite, version 14, 1996) was employed to backup all DIS PDU traffic on the network. To initiate a radio communication, team members pressed a PI Engineering X-Keys foot pedal which activated the ModIOS software. During the experiment, teammates communicated on the same radio frequency (i.e., communications made by one team member were received by all team members). This was done deliberately to simulate the saturated communication channels encountered by personnel in modern military environments. To communicate effectively, participants had to adopt communication strategies that emphasized accuracy and brevity.

To enforce the use of headsets and microphones for oral communication between participants, and to simulate the auditory environment experienced by personnel aboard C2 platforms such as the E-3 Sentry, background noise was generated in the laboratory during experimental trials. A Bruel & Kjaer Noise Generator (Type 1405) was employed to produce a 50 kHz pink noise. The BNC output of the noise generator was converted to left and right component plugs using a cable adapter and fed into a NAD 2100 Monitor Series Power Amplifier. The amplifier was connected to two Magnepan Magneplanar SMGa 4 ohm loudspeakers. The background pink noise produced by this system was approximately 55 dB.

2.2.4.1. Synthetic Task Environment

The simulated environment utilized in this experiment was Aptima, Inc.'s Distributed Dynamic Decision-making (DDD) software (version 3.0; MacMillan, Entin, Hess, & Paley, 2004). DDD is a tool for creating scriptable, low-to-moderate fidelity, human-in-the-loop multi-participant simulations. Its software architecture is designed using a client-server model, written in C, and is Linux based. DDD has successfully been used to simulate team command and control tasks and to study realistic and complex team processes in a variety of military and civilian research projects (MacMillan et al., 2004). The DDD was employed in this experiment to create a set of air defense simulations conveyed to participants through a tactical display.

The task utilized in this experiment required five-person teams to work together to complete a simulated air defense command and control (C2) task. This task has been used in several previous experiments examining collaborative tool usage in military settings and has been demonstrated to be sensitive to experimental manipulations (e.g., Finomore, Knott, Nelson, Galster, & Bolia, 2007). The scenario required a team comprised of two weapons directors (WDs), two sweep operators, and one tanker operator; these positions differed in their roles and capabilities. Weapons directors were responsible for matching friendly fighters with appropriate enemy targets, scheduling fighters for refueling and resupply, and communicating their action plans with other team members. Strike and tanker operators maneuvered team assets as instructed, engaged enemy targets, and provided pertinent information to teammates concerning asset resources (i.e., weapon and fuel status). In this experiment, participants were always assigned to the WD positions and confederates to the sweep and tanker positions. As such, participants had primary decision making and leadership responsibility. Confederates, on the other hand, were instructed to carry out the orders given to them by the participant WDs as accurately and quickly as possible without providing advice or strategy concerning task execution.

The experimental simulation was presented to team members by means of the DDD tactical display. The tactical display included representations of the area of operations and of friendly and enemy assets, which were depicted using unique, non-overlapping symbols. The display also exhibited the movements of aircraft within the battle space and provided information about them such as speed, heading, weapons and sensor ranges, fuel, and weapons status.

Depicted in Figure 2 is an example of the WDs' tactical display. The display provided WDs a global picture of the simulated battlespace, comprising all team assets and enemy aircraft. However, WDs were not afforded direct control of team assets. Rather, they used the DDD display to monitor the simulation and used communications software to issue directives to the sweep and tanker operators. Strike and tanker operators used the DDD to maneuver team assets and retrieve information, but the locations of enemy aircraft were hidden until they came within a short distance of team fighters. Therefore, confederate operators had limited awareness of the tactical situation during a trial and had to rely on the participant WDs to vector them to targets.

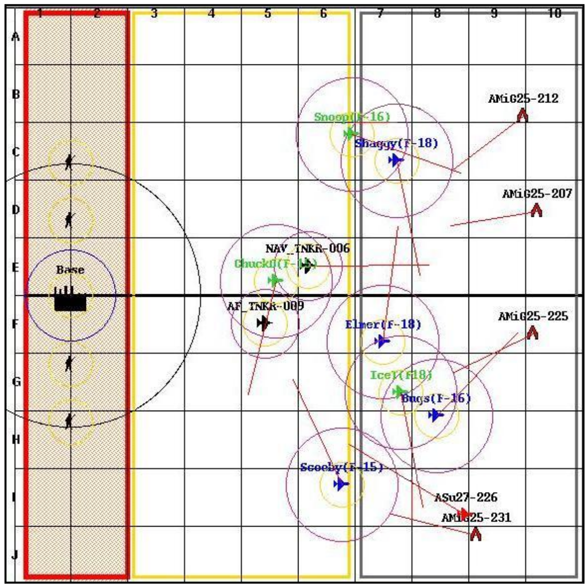

Figure 1. An example of the DDD tactical display. Team assets are represented as blue, green, and black symbols, and enemy assets as red symbols. Participants were charged with preserving team assets and preventing enemy aircraft from entering the yellow and red "friendly" zones.

Within the simulation, the two WDs managed separate assets and geographical areas of responsibility (AORs). Participants were instructed that they were jointly accountable for performing the air defense task, but that each would be assigned primary responsibility for the northern or southern sector of the battlespace (the division between AORs was indicated by a solid black horizontal line). Team assets were represented as blue, green, or black symbols (stylized aircraft icons), and each was labeled with a fixed callsign (e.g., "Elmer") and platform designation (e.g., F-16). These assets were color-coded such that the WD responsible for the northern AOR (the "Green WD") controlled green assets, while the southern AOR WD (the "Blue WD") controlled blue assets. Tanker aircraft, represented in black, were team assets and had to be shared by both WDs. Although each WD's assets operated primarily within their AOR, participants were instructed that they were free to cross AOR boundaries to provide assistance or enact team strategies. In addition, the battlespace featured gray, yellow, and red "engagement" zones. Participants were instructed to prosecute enemy aircraft in the gray zone and to prevent them from entering friendly airspace (i.e., the yellow and red zones).

This experiment featured two classes of hostile targets (MiG-25, Su-27), which were differentiated by their on-screen representations and their speed of movement. The majority of enemy targets in each scenario were MiGs, which were slightly slower than WD fighter assets and were represented in the simulation by a red, inverted "V." Su-27 targets, on the other hand, were slightly faster than WD fighter assets, necessitating frontal interception by team assets, and were represented by a red aircraft icon. The number of enemy targets present throughout each trial was deliberately controlled. Each trial featured six Su-27 aircraft, which appeared at random intervals in the scenario. Conversely, each time a MiG was intercepted and destroyed, a new one would enter the scenario to replace it. This generation rate of enemy aircraft ensured a relatively constant level of task load throughout each trial. All enemy targets entered the scenario from the right side of the display (in the gray zone), and proceeded on a random path to the left side of the display (the red zone). As they moved through the simulated battlespace, enemy aircraft could attack and destroy the team's fighter and tanker assets, an Air Force base, and four ground-based infantry units positioned in the red zone.

WDs' primary duties included relaying tactical information to their assets, directing assets to intercept hostile targets, and coordinating aerial refueling between fighter and tanker assets. To do this effectively required WDs to perceive the capabilities and limitations of their operational environment. Within the simulation, three classes of friendly fighter assets (F-15, F-16, F-18) were employed. F-15 and F-16 assets were equipped with two missiles and could only target enemy MiGs. F-18 assets were outfitted with four missiles, two for attacking MiGs and two for attacking Su-27s. At the beginning of a scenario, each WD was responsible for one F-15, one F-16, and two F-18s. WDs also had access to two tanker assets used for airborne refueling and weapon restocking (this is a departure from real world capabilities, in that tankers cannot re-arm other aircraft). The Air Force tanker was able to refuel and restock F-15 and F-16 assets, while the Navy tanker was only able to refuel and restock F-18 assets. In addition, participants could refuel and restock any fighter asset at an Air Force base, located in the red zone. Experimental and practice trials in this experiment were ten minutes in duration. At the start of each trial, all fighter assets began with a randomized fuel level below their maximum capacity of eight minutes. Fighter assets' fuel reserves depleted at a constant rate requiring refueling at least once during each trial.

2.2.4.2. DDD_Results Application

The *DDD_Results* (2006) application is custom software created by the AFRL, written in Visual Basic (VB) using Microsoft's .NET framework. This software created a detailed log of the events that occurred during each experimental trial and generated feedback for participants in the form of a "team score." This score reflected how well the team achieved the scenario goals. This score was scaled so it could range from 0-100; a score of 0 indicated that the team did not meet any of the goals of the scenario, and a score of 100 indicated that the team met all of the goals perfectly. The team score was generated based on three equally weighted performance factors: a) prevention of enemy incursions into friendly airspace, b) preservation of team assets, and c) protection of friendly ground forces in the red zone (the air base and infantry units).

In generating the log file and team score, DDD_Results executed several steps. First, a File Transfer Protocol (FTP) command was sent to the Red Hat Linux server to retrieve the history files generated by DDD at the completion of a trial. Using Python (version 2.2.3; 2003), an open source programming language designed to provide code readability, the DDD history files were then converted into Comma Separated Value (CSV) files. Next a Microsoft Excel 2003 file was opened, which initiated an embedded custom macro to extract the trial data from the CSV files and populate the cells of the Excel worksheet. Finally, the team score was calculated and displayed to participants in a pop-up window.

2.2.4.3. Auditory Monitoring Task

In addition to the primary air defense task, participant WDs performed a secondary auditory monitoring task. The task employed in this experiment was adapted from Bolia, Nelson, Ericson, and Simpson (2000) and was designed to assess speech comprehension in a multi-talker environment. In the current experiment, the task was used to further simulate the complex communication demands experienced by personnel in military environments. Task stimuli consisted of a call sign, a color, and a number embedded in a carrier phrase (e.g., "Ready Baron go to red six now," "Ready Baron go to blue eight now"). Participants listened for messages addressed to the call sign "Baron" and responded by activating the button corresponding to the color and number combination indicated from a larger response matrix (the matrix is presented in Figure 3). To increase the difficulty of the task, a second, similar message always accompanied the target message. This distracter message incorporated the same elements as the target message, but was addressed to a distracter call sign.

Target and distracter messages were drawn from the Bolia et al. (2000) speech corpus, which includes eight call signs (‘‘Arrow,’’ ‘‘Baron,’’ ‘‘Charlie,’’ ‘‘Eagle,’’ ‘‘Hopper,’’ ‘‘Laker,’’ ‘‘Ringo,’’ ‘‘Tiger’’), four colors (‘‘blue,’’ ‘‘green,’’ ‘‘red,’’ ‘‘white’’), and the numbers one through eight. The 256 phrase combinations of these elements were recorded by each of eight speakers, four men and four women, for a total of 2048 phrases in the corpus. Each recorded message is approximately 1.5 seconds in duration. In this experiment, target and distracter messages were presented asynchronously to participants with a 10 ms delay between the start of each. The serial order of messages (target-distracter, distracter-target) was counterbalanced across presentations. Messages were broadcast to WDs every 30 seconds (different target and distracter messages were sent to each participant), for a total of 20 target messages per trial. Target-distracter couplings were organized for maximal disparity, such that paired speakers were always of opposite gender (one man and one woman), and colors and numbers were not permitted to overlap between messages (i.e., if the target message was “blue seven,” the distracter message could not include “blue” or “seven”). Additionally, messages were counterbalanced across experimental trials so that all colors and numbers were presented as targets and distracters approximately equally.

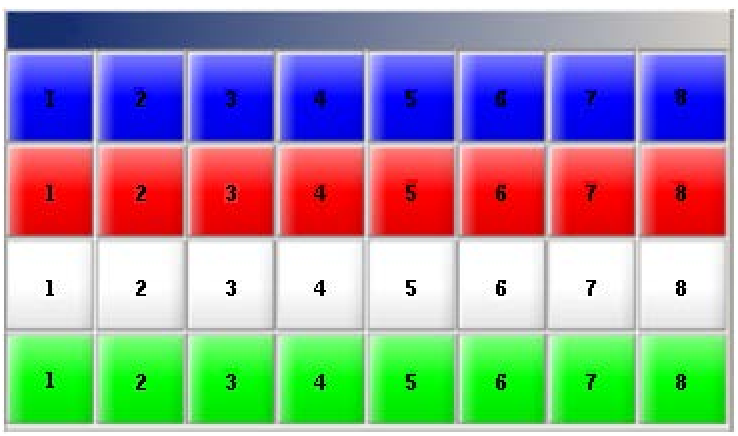

Figure 2. The auditory monitoring task response matrix. Participants listened for a color and number combination, such as “white five,” and responded by pressing the matching button from the matrix.

2.2.4.4. Dynamic Real-time Animated Whiteboard (DRAW)

The *dynamic real-time animated whiteboard* (DRAW; 2006) application is custom software created by the AFRL, and is written in VB using Microsoft’s .NET framework in conjunction with SQL Server for configuration and data storage. Based on the suggestions of Bolstad and Endsley (2005), DRAW was created to be a domain-specific graphical collaboration tool tailored specifically for military applications. It allows users to quickly and easily communicate information, particularly spatial information, using a lexicon of pre-programmed “drag-and-drop” symbols, with the intent of providing an alternative, but complementary, communication medium to auditory (radio) channels in military environments. The intent was to provide a means to expeditiously convey critical decisions and command intent across the chain of command, allowing users to maintain a high level of situation awareness while performing their current and future duties. DRAW is “dynamic” in that it can be used to add tactical and iconographic information to any application, "real-time" in that command directives may be rapidly distributed to all users, and "animated" in that annotations appear in a transparent layer over the target application on a virtual "whiteboard" surface (Figure 4).

In this study, DRAW operated conjointly with a second custom application, *ScreenCapture* (2003). ScreenCapture was also written in VB .NET, and was designed by the AFRL to automatically record the user's computer screen and save that image as a .jpg image file on a polled interval (one image per second). The last-captured screen image was then imported into DRAW for annotation. This approach provided a benefit over currently available commercial software by automatically importing an annotatable image. Other commercial white boarding applications allow users to import, annotate, and share images, but they require additional, manually-input commands from users to accomplish the procedure, making them less suitable for high-tempo command and control environments.

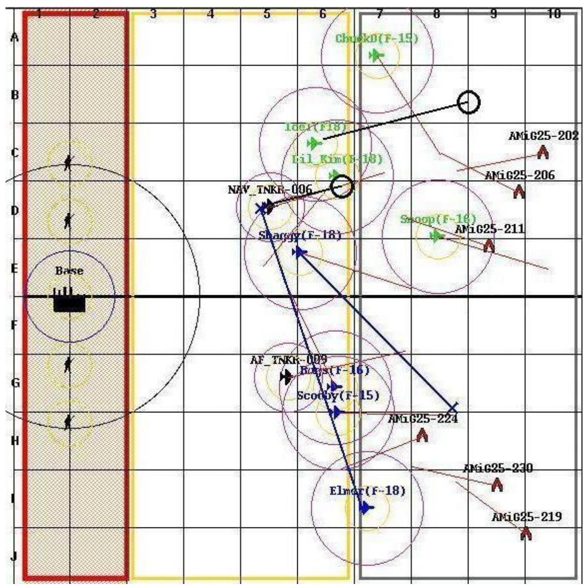

Figure 3. An example of the DDD tactical display annotated with user-created DRAW commands. DRAW allows users to add annotation to other software applications and share the generated images with other users. In this figure, the DRAW marks (which appear as black lines terminating in "X" or "O") denote movement directives for several team assets.

2.2.4.5. Resource Display

The Resource Display (2006) software, created by the AFRL, was written using Microsoft's .NET framework utilizing the VB software language. This software was designed to display team assets' weapon and fuel state information to team members. The software utility connected to the DDD simulation using a TCP/IP socket connection and acted as an additional DDD client. As depicted in Figure 5, the resource display extracted relevant asset information from the DDD simulation and displayed it for participants in a digital format.

ChuckD	**Fuel:**	**04:48**	**MiG:**	**1**	**Su:**	**---**
IceT	**Fuel:**	**05:19**	**MiG:**	**1**	**Su:**	**0**
Lil_Kim	**Fuel:**	**06:14**	**MiG:**	**1**	**Su:**	**2**
Shaggy	**Fuel:**	**03:58**	**MiG:**	**0**	**Su:**	**1**
Bugs	**Fuel:**	**03:19**	**MiG:**	**0**	**Su:**	**---**
Elmer	**Fuel:**	**06:54**	**MiG:**	**1**	**Su:**	**2**
Scooby	**Fuel:**	**06:32**	**MiG:**	**0**	**Su:**	**---**
Shaggy	**Fuel:**	**03:58**	**MiG:**	**0**	**Su:**	**1**

Figure 4. An example of the resource display's digital readout.

Figure 6 illustrates the spatial arrangement of the DDD tactical display, the DRAW digital whiteboard, the auditory monitoring task matrix, and the resource display as they were arrayed in the Windows environment.

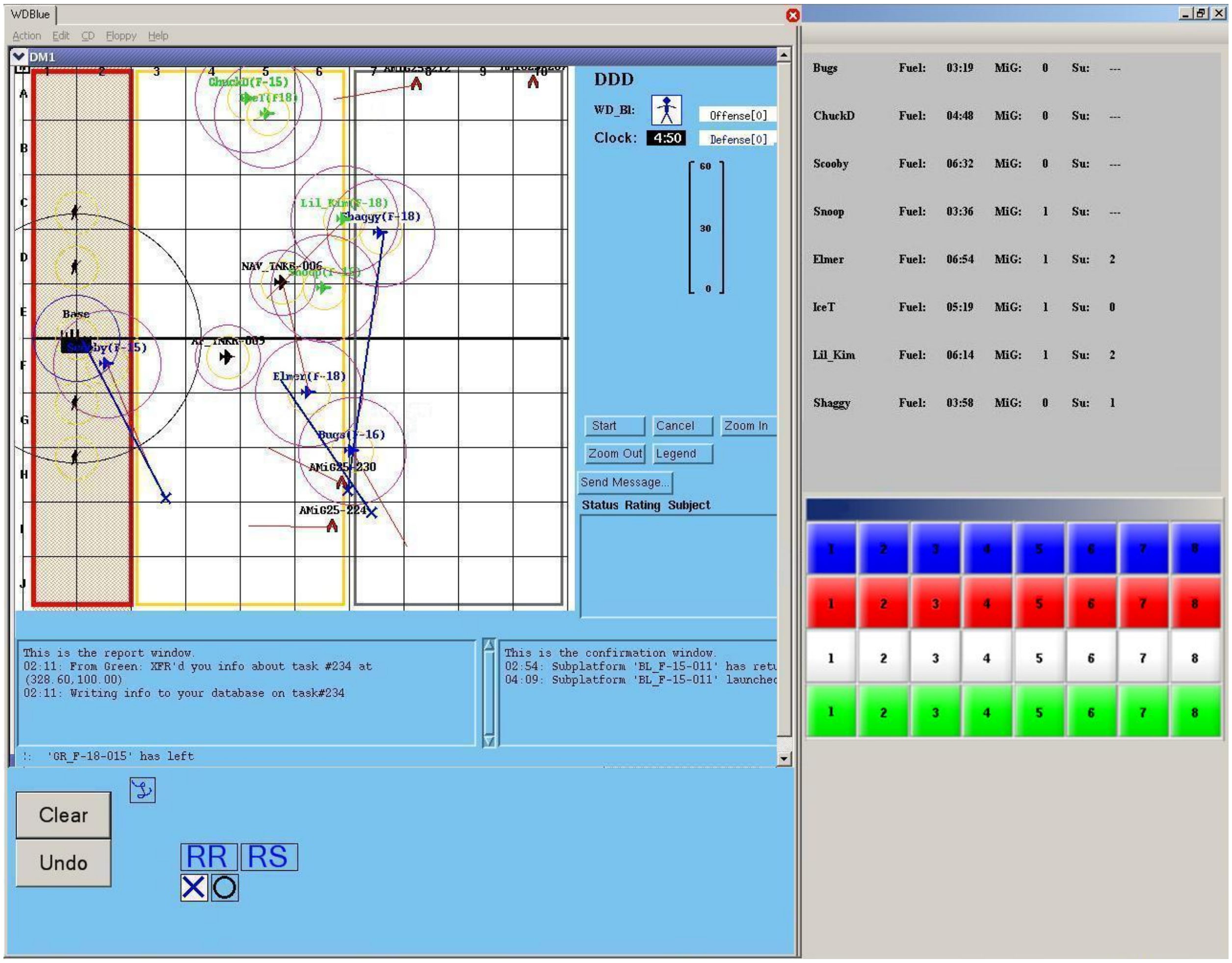

Figure 5. An example of the spatial configuration of the Windows environment in this experiment. Depicted are the DRAW software window, which overlaid the DDD tactical display (left half of the screen), the resource display (upper right), and the auditory monitoring task response matrix (lower right).

2.2.4.6. Microsoft Virtual PC and Synergy

Microsoft Virtual PC 2004 (MVPC; 2004) was used to host a virtual machine running the Red Hat Linux 9.0 (RH9; 2004) operating system required for the synthetic task environment software employed in this experiment. A virtual machine is a software implementation of standardized computer hardware that enables a user to execute applications in a fashion similar to that provided by a computer actually equipped with that hardware. MVPC was necessary to enable the Linux based synthetic task environment software to co-exist with other applications running in the Windows environment. This configuration allowed the participants and experimenters a homogeneous software environment in which to operate.

MVPC does allow a workstation's mouse and keyboard to operate within the virtual machine by clicking inside the MVPC window. However, once activated in this fashion, MVPC requires users to depress the right ALT button on the keyboard to release control of the mouse and keyboard and return to interaction with applications outside MVPC (i.e., other applications in the Windows environment).

Synergy (version 1.2.7; Schoeneman, 2002) was employed to share mouse and keyboard inputs between the RH9 virtual machine and the host Windows computer. Synergy is an open source software application that enables a user to share a single mouse and keyboard across multiple computers when each computer has its own display and operating system. Synergy was configured in this experiment to facilitate the transition of mouse and keyboard inputs between the Windows and Linux environments. Synergy was only implemented on the confederate workstations, as they were the only users required to interact with both the virtual machine and Windows.

2.2.4.7. DDD_Console

The *DDD_Console* application (2006) is custom-built software developed by the AFRL, written in the Visual Basic software language using Microsoft's .NET framework. This software allowed experimenters to select the experimental conditions for each trial, start and stop trials in the DDD environment, and it automatically generated a time-stamped log of those events in a SQL database. When initiating a trial, the DDD_Console software communicated with the Red Hat Linux server, via a telnet command, and instructed it to begin the DDD simulation software. Additional telnet commands were then sent to participants' computers to activate each as a DDD client, and following completion of the trial, to terminate the DDD software.

In conjunction with the DDD_Console, Microsoft's PsTools Suite (version 1.6, 2006) enabled experimenters to initiate and terminate software applications in the experiment. The Suite contains a number of command-line tools designed to assist in the management of local and remote systems. Specifically, the *PsExec* tool was used to start applications, and *PsKill* stopped them.

In addition, PuTTY (version 0.59; Tatham, 2007) was used by the DDD_Console as a bridge between the Windows and Linux operating systems. Specifically, PuTTY provided Telnet and Secure Shell (SSH) access from Windows to Linux through a network connection. The DDD_Console software stored Telnet commands in a batch file, and implemented them by PuTTy Link command (Plink), which instructed PuTTY to send the commands in the batch file to the DDD server.

2.2.4.8. Morae

TechSmith Corporation's Morae (version 1.3, 2005) application used a web camera to record video of each workstation's display and user (participant or confederate) at a rate of three frames per second. Morae was also configured to record each user's radio communications and all mouse, keyboard, and window events. Across workstations, Morae was configured to start recording at the same time in order to provide temporal synchronicity for post-experiment analysis and playback.

2.2.5. Questionnaires

Participant WDs completed several questionnaires during this experiment; experimental confederates were not required to complete these measures. All questionnaires employed in this experiment were administered to participants in an electronic format (i.e., the paper-and-pencil version of each was recreated as a graphical user interface in the Windows PC environment). The Subject Survey System (SSS; 2003), created by the AFRL and written in the Java 2 programming language (version 1.4.2), was utilized to distribute questionnaires to participants and record their responses. The SSS uses a client-server architecture. Once the software is initiated by the experimenter, the server queries a Microsoft Access database for a list of configuration specifications of relevant computers and a list of the questionnaires that could be submitted to each user. The SSS server then parses its list of computers and, using PSTools, launches the clients using a batch script stored separately on each computer. When the client initializes, it receives the name of the computer hosting the SSS server, allowing it to connect to and poll the server for further configuration information including the exact questionnaires to display and experimental details (e.g., participant identification number, trial number, etc.) used to identify the data. Following completion of all questionnaires, each SSS client connects to the SSS server via Java's Remote Method Invocation (RMI) and saves questionnaire responses to a Microsoft Access 2003 database.

2.2.5.1. NASA-Task Load Index

Following completion of each experimental trial, WDs completed the NASA-Task Load Index (TLX; Hart & Staveland, 1988), a standard measure of workload that is widely used in human performance research (Wickens & Hollands, 2000). The NASA-TLX provides a global index of task workload on a scale of 0 to 100 and identifies the relative contributions of six sources of workload: mental demand, temporal demand, physical demand, performance, effort, and frustration.

2.2.5.2. Modified-TLX

After each trial, WDs also completed a version of the Modified-TLX (M-TLX; Pharmer, Cropper, McKneely, & Williams, 2004). The M-TLX is an unvalidated measure designed to assess potential drivers of workload in team settings, and is comprised of five subscales: communication demand, monitoring demand, control demand, coordination demand, and leadership demand. It was included in this experiment to address suggestions made by Bowers, Braun, and Morgan (1997), who have argued that the NASA-TLX may not adequately capture sources of workload present during team tasks. In this experiment, participants rated each subscale from 0 to 20 on three dimensions: degree of demand (low/high), difficulty performing subscale-related behaviors (easy/hard), and frequency of subscale-related behaviors (infrequent/continuous). Subscale scores were calculated as the sum of the three dimensional scores; consequently, subscale ratings could range from 0 to 60. In addition, a global M-TLX workload rating was calculated by computing the mean of the five subscales.

2.2.5.3. Modified-MRQ

Following each two-trial communication condition block, WDs completed a version of the Multiple Resources Questionnaire (MRQ; Boles & Adair, 2001). The Modified-MRQ (M-MRQ; Finomore et al., 2006) asks participants to rate the extent to which a task they have performed utilized 17 resource dimensions drawn from Wickens' multiple resource theory (Wickens & Hollands, 2000). The resource dimensions of the MRQ are presented below in Table 2. Research using the M-MRQ indicates that it possesses greater sensitivity than the standard MRQ without modifying its diagnostic profile, and that it may be useful in identifying sources of task workload that are not represented in the NASA-TLX (Finomore et al, 2006). Items on the M-MRQ are scored from 0 (no usage) to 100 (extreme usage).

Table 2. The 17 M-MRQ Resource Dimensions.

Subscale	Abbreviation	Subscale	Abbreviation
Auditory emotional	AE	Spatial emergent	SE
Auditory linguistic	AL	Spatial positional	SP
Facial figural	FF	Spatial quantitative	SQ
Facial motive	FM	Tactile figural	TF
Manual process	MP	Visual lexical	VL
Short-term memory	STM	Visual phonetic	VP
Spatial attentive	SA	Visual temporal	VT
Spatial categorical	SC	Vocal process	V
Spatial concentrative	S		

2.2.6. Procedure

As mentioned previously, several team roles employed in this experiment were performed by confederates. Before acting in this capacity, all confederates completed a behavioral training session which included information concerning their responsibilities in the simulated air defense task and appropriate conduct during the experiment. Specifically, confederates were told to regularly update the two participant WDs concerning their assets' fuel and weapon states and to follow the orders given to them by the WDs without providing specific strategies for task execution. Following the behavioral training session, each confederate was assigned to a specific team role (blue sweep, green sweep, or tanker operator) and received twelve hours of practice in that role. It is important to note that, though confederate performance was integral to overall team performance in the simulated air defense task, the authors were primarily interested in the performance and subjective responses of the participant WDs.

Prior to experimental data collection, all participant WDs completed a four-hour training session. During this time, they received training on the simulation, the radio software, DRAW, and the resource display. Additionally, participants were trained on and practiced communication brevity for oral communications. Brevity training was critical to minimize the number of irrelevant, unnecessarily lengthy, or confusing communications that teams might make.

Participants were informed that the purpose of the study was to evaluate how teams used communication technology to work together and that they would be playing a computer game that required teamwork to meet the game's objectives. They were further instructed that the performance of the team would be scored following each trial for how well they had met their objectives and followed the rules of the simulation (as described above).

Participant WDs were then administered a short review test designed to assess their recollection of the previously presented training information. They were required to answer all items on the review correctly before continuing with the training session (participants were permitted to re-take the test if they answered any items incorrectly). Following the test, teams completed 11 practice trials, allowing them to further familiarize themselves with the task and collaboration tools employed in the experiment.

Teams returned the next day for the experimental session. Upon arrival in the laboratory, they were assigned an order of presentation of the experimentally manipulated factors. The experimental schedule of conditions was counterbalanced across teams to control order effects. During the experiment, teams completed sixteen trials, eight experimental trials and eight practice trials. Trials were presented in a block fashion, with each block consisting of four trials in the same communication and resource display conditions. The first two trials of each block were practice trials, and did not include the auditory monitoring task. The remaining trials of each block were experimental trials, and did feature the monitoring task.

Participants were given one 20-minute rest period after they had completed four experimental trials. The experimental session was completed in approximately four hours. During each trial, the simulation events (e.g., occurrences and outcome of attacks, refuelling events, etc.) were recorded in data logs for later analysis. In addition, Morae recorded all video and radio communications during each trial.

2.3. Results

2.3.1. Team Performance

During each experimental trial, software recorded several indices of team performance including the team score, the number of enemy aircraft intercepted, the total time required to prosecute an enemy aircraft (i.e., the time from an enemy aircraft's appearance in the simulation until it was intercepted, in seconds), the percentage of enemy aircraft that successfully penetrated friendly airspace, and the number of team assets lost. Displayed in Table 3 are the means for each performance variable in each condition.

Table 3. Mean team performance across several task indices as a function of team communication and resource display conditions.

	Performance Variables				
Trial Condition	Team Score	Enemy Aircraft Intercepted	Time to Prosecute	Airspace Penetration	Team Assets Lost
Standard Communication					
RD Absent	72.59 (5.50)	26.38 (1.41)	121.72 (4.89)	37.93 (4.57)	3.38 (.94)
RD Present	73.62 (3.84)	27.44 (1.26)	116.49 (4.31)	36.17 (2.58)	3.25 (.78)
Augmented Communication					
RD Absent	82.25 (3.18)	29.50 (1.00)	107.84 (1.45)	30.15 (2.17)	1.75 (.59)
RD Present	74.25 (4.52)	28.38 (1.15)	111.50 (1.75)	30.77 (2.92)	3.44 (.87)

Note. RD = Resource display. Values in parentheses are standard errors.

To examine the effects of the experimental manipulations on team performance, the mean was calculated for each team on each variable. These values were then tested for statistically significant differences using separate 2 (team communication) × 2 (resource display) repeated measures analyses of variance (ANOVAs).

The results of these analyses revealed statistically significant differences between conditions when measuring time required to prosecute an enemy aircraft. Analysis of this performance measure indicated a statistically significant main effect of *team communication* condition, F (1, 7) = 5.13, p < .05. Teams were able to intercept enemy targets more quickly when they had access to the virtual whiteboard in the augmented communication condition compared to when they did not, perhaps due to improved spatial awareness provided by the DRAW's pictorial representations of the simulated battlespace. No significant differences were found between conditions for any other performance variables (all main effects and interactions p > .05).

2.3.2. Auditory Monitoring Task Performance

The CRM program recorded the number of signals responded to and the number of correct responses each participant made in each trial. However, due to a computer error the response data of three participants was lost and could not be recovered. The mean number of responses and correct responses were computed for the remaining 13 participants and analyzed for statistically significant difference between conditions using separate 2 (team communication) × 2 (resource display) repeated measures ANOVAs.

Across conditions, response rate to the auditory monitoring task was relatively low (participants responded to approximately 60% of the signals). The results of the analysis for the number of signals responded to indicated a statistically significant main effect of *team communication* condition, F (1, 12) = 5.33, p < .05. Participants made more responses to the monitoring task in the augmented communication condition (M = 12.33, SE = 1.18) compared to the standard condition (M = 11.25, SE = 1.31). No other sources of variance in the analysis were significant (all p > .05).

For the number of correct responses, a statistically significant interaction between *team communication* and *resource display* conditions was detected, F (1, 12) = 4.89, p < .05. Follow-up simple main effects paired-sample t-tests for each communication condition indicated that, in the augmented communication condition, participants made more correct responses to the task on trials when they did not have access to the resource display compared to trials when they did, t (12) = 2.95, p < .05. However, no such difference was found between resource display conditions in the standard communication condition (p > .05). This relationship is illustrated in Figure 7. In these and subsequently reported post hoc analyses, the Dunn-Sidak alpha correction was applied to control Type I error rates (Kirk, 1995). The observed differences in secondary task performance with access to the virtual whiteboard suggest that DRAW allowed participants to reduce radio communication saturation, resulting in improved auditory task comprehension. This effect seems to be somewhat reduced by the addition of the resource display, perhaps because of the need for participants to divide attention across the tactical and resource displays.

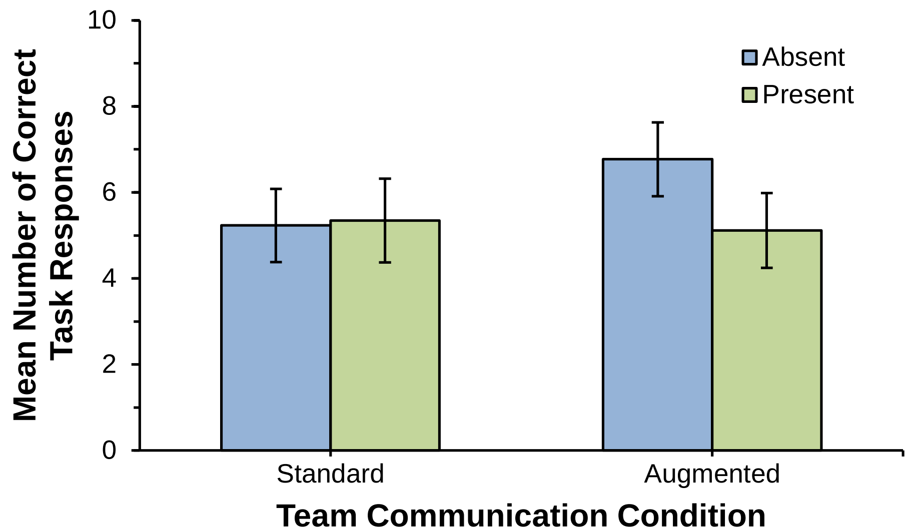

Figure 6. Mean numbers of correct auditory monitoring task responses as a function of team communication and resource display conditions. Error bars are standard errors.

2.3.3. Team Communication

Following the completion of experimental data collection, audio recordings and DRAW logs of the communications between teammates were compiled and examined. Across trials, teams sent an average of 122.40 radio messages per trial. In addition, when the virtual whiteboard was available, teams sent an average of 77.80 DRAW messages per trial. As a manipulation check, the mean numbers of DRAW marks sent per trial were tested against a value of zero using a one-sample *t*-test to establish that teams were using the tool. The results of this analysis indicated that participants were communicating at a rate greater than zero using DRAW marks, $t\ (7) = 13.62$, $p < .05$.

2.3.3.1. Virtual Whiteboard Communication

To examine the number of virtual whiteboard communications sent for potential differences related to the availability of the resource display, a paired-samples *t*-test was computed comparing absent and present trials in the augmented communication condition. The results of the analysis indicated that teams sent approximately the same number of DRAW communications in each resource display condition, $t\ (7) = 1.42$, $p > .05$.

2.3.3.2. Radio Communication

Using the XML summary created by WCAS, the frequencies and durations of team communications during each trial were computed. Frequency was calculated by summing the number of communications, and duration by summing the total length of radio communications during a trial (each measure was calculated irrespective of speaker). Mean values were then calculated for each team and experimental condition; these values were tested for statistically significant differences between conditions using separate 2 (team communication) × 2 (resource display) repeated measures ANOVAs. For the frequency of radio communications, statistically significant main effects were found for the *team communication*, $F\ (1, 7) = 68.09$, $p < .05$, and *resource display* conditions, $F\ (1, 7) = 9.86$, $p < .05$. No other sources of variance in the analysis were statistically significant ($p > .05$). As is depicted in Figure 8, participants made significantly fewer radio communications when they had access to the virtual whiteboard and when they had access to the resource display.

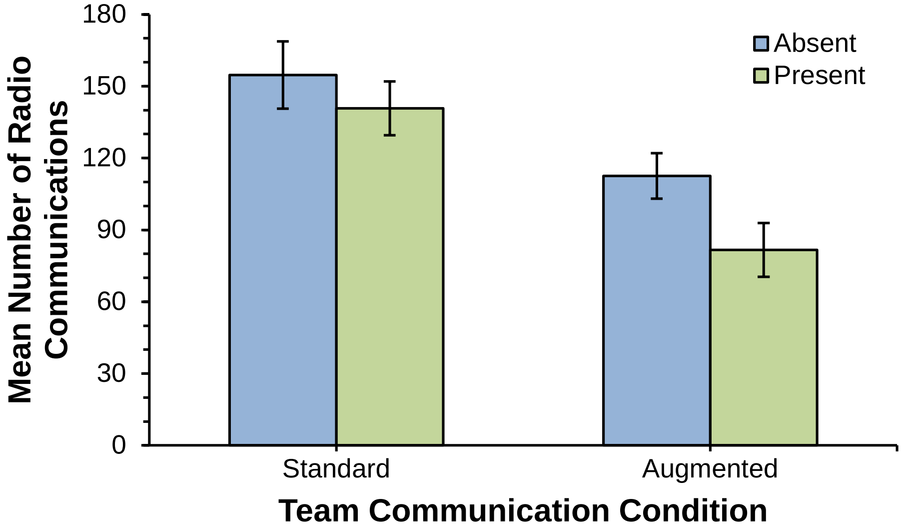

Figure 7. Mean number of radio communications as a function of team communication and resource display conditions. Error bars are standard errors.

For the duration of radio communications, a statistically significant main effect of *team communication* condition was detected, $F\ (1, 7) = 79.07$, $p < .05$. No other sources of variance in the analysis were statistically significant ($p > .05$). The average duration of all radio communication during a trial was approximately 75% greater in the standard communication condition (M = 445.39 s, SE = 21.27 s) compared to the augmented condition (M = 257.01 s, SE = 21.73 s). Overall, the observed reductions in frequency and duration of radio communication in the augmented condition suggest that teams were able to successfully transition communication from the saturated radio channel to the virtual whiteboard.

2.3.3.3. Radio Communication Content Analysis

Following completion of experimental data collection, all radio communications between participants were hand transcribed, resulting in a total of 7,992 transcribed communications. A content analysis of these communications was then conducted. The categorization scheme employed was specifically developed for this experiment. Short descriptors of the eight categories employed appear in Table 4.

Table 4. Short descriptions of the eight categories used for content analysis in this experiment.

Category	Description	Example Statements
Clarification / Confirmation	Statements about either complying with or clarifying an order or request.	"*Copy that.*" "*Say again?*"
Coordinate	Statements which reflect planning, back-up behavior, or assisting teammates, but which were not directives for action.	"*I need some help down here.*" "*Can I move the navy tanker?*"
Directive – Maneuver / Attack	Statements concerning maneuvering fighters or tankers. Most maneuver directives were for the purpose of positioning a fighter to intercept an enemy aircraft, but this category also included repositioning assets to avoid enemy aircraft as well.	"*Intercept MiG 227 at G6.*" "*Tanker relocate to E4.*"
Directive – Resupply	Statements tasking assets for refueling or resupply. It included any maneuver statements that were clearly for the purposes of refueling fighters or resupplying their weapon loads.	”*Refuel at navy tanker.*" "*Restock and refuel your fighters.*"
Resource Status Request	Questions concerning asset fuel or weapons loadings.	"*Who has low fuel?*" "*How many arms remaining?*"
Resource Status Update	Statements that provide information about fuel or weapon loadings.	"*I've got 1 minute of fuel left.*" "*No arms remaining.*"
Situation Update	Statements or questions concerning scenario events and developments. These communications were often intended to provide awareness to team members about significant events or an update to a previous directive.	"*Did we lose a fighter?*" "*There are still two MiGs at I5.*"
Social / Emotive	Statements which reflected emotion, social interaction, or performance feedback, but were not directly related to performing the task.	"*Good job!*" "*Did you see the game last night?*"

In conducting the content analysis, two judges independently classified each transcribed radio communication as an instance of a single category. Interrater reliability of the judges, assessed by the proportion of overall agreement (Uebersax, 2000) and Cohen's kappa (Cohen, 1960), was deemed by the authors to be sufficient (proportion of overall agreement = .93; Cohen's kappa = .90, $p < .05$). The percentage of radio communications in each category for each experimental condition is presented in Table 5. As can be observed in the table, access to the resource display resulted in relatively substantial decreases in the percentage of radio communications classified as *resource status – update* and *resource status – request*, which is consistent with the information conveyed by the display, and an increase in *social* communications. Access to the virtual whiteboard in the augmented communication condition resulted in decrements in the percentage of radio communications classified as *directive – attack* and *directive – resupply*, which is consistent with the types of communication the DRAW was designed to convey, and an increase in the number of *clarification / confirmation* and *social* communications.

Table 5. Percentage of radio communications by category as a function of team communication and resource display conditions.

	Absent		Present		
Category[a]	Standard	Augmented	Standard	Augmented	Percentage of Total[b]
Clarification / Confirmation	30.56	39.32	32.81	43.78	35.84
Directive – Maneuver / Attack	24.94	10.91	35.01	7.26	21.20
Situation Update	11.55	13.97	11.6	18.36	13.43
Resource Status – Update	13.44	18.05	2.74	6.12	10.04
Directive – Resupply	9.11	5.35	13.24	2.29	8.15
Social	3.13	2.91	4.26	20.71	6.68
Resource Status – Request	6.76	9.13	0.17	0.34	4.19
Coordinate	0.51	0.36	0.17	1.14	.49

[a]Categories are presented in their order of predominance, from largest to smallest, in the complete 7,992 item data set.

[b]Indicates the prevalence of communications in each category from the complete data set, collapsed across experimental conditions to facilitate cross-condition comparisons.

2.3.4. Subjective Workload Measures

To test the effects of the experimental conditions on participants' evaluations of task workload, mean ratings for the six NASA-TLX subscales, the five M-TLX subscales, and the 17 MRQ subscales were computed for each participant. Workload ratings from each measure were then tested for statistical significance by means of separate 2 (team communication) × 2 (resource display) × 6, 5, or 17 (TLX, M-TLX, and MRQ subscales, respectively) repeated measures ANOVAs. Following the suggestion of Muller and Barton (1989), in these and all subsequently reported analyses involving repeated measures with more than two levels of the factor, the Box/Geisser-Greenhouse epsilon correction was employed to adjust the ANOVA degrees of freedom, ameliorating violations of the sphericity assumption.

2.3.4.1. NASA-TLX Workload

The mean TLX workload rating, computed across subscales, reported in this experiment was 42.73 (SE = 1.41). This value is near the midpoint of the scale, indicating that participants found the ABM task to be moderately to highly demanding.

The ANOVA analysis of the TLX workload ratings indicated a statistically significant main effect of *TLX subscale*, F (3.29, 49.41) = 14.42, $p < .05$, and a statistically significant interaction between *team communication* and *resource display* conditions, F (1, 15) = 4.57, $p < .05$. No other sources of variance in the analysis were significant (all $p > .05$). As is depicted in Figure 9, the mental demand, temporal demand, and effort associated with the task appear to be drivers of participants' workload estimates.

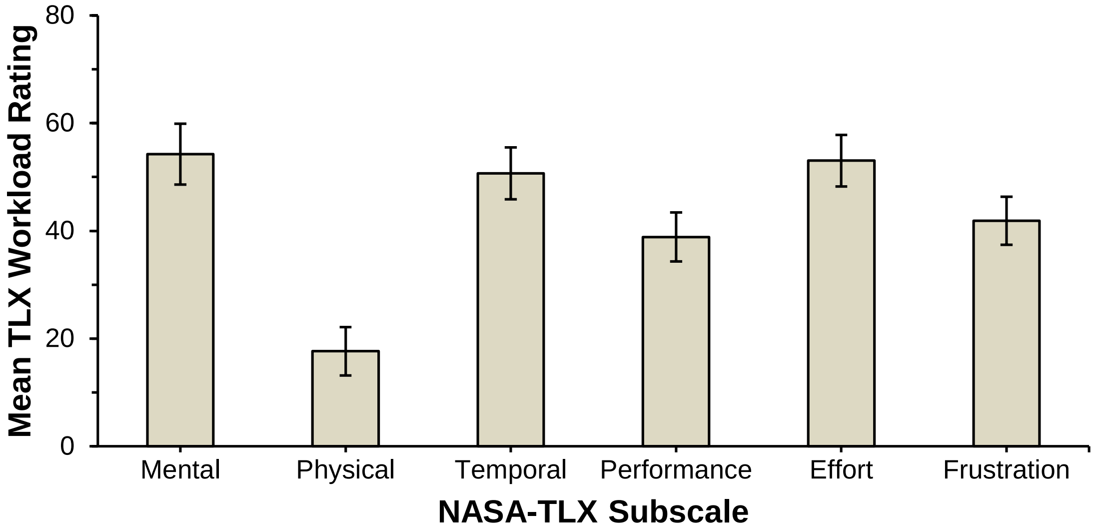

Figure 8. Mean workload ratings for each of the NASA-TLX subscales. Error bars are standard errors.

In exploring the *team communication* by *resource display* interaction, follow-up simple main effects paired-sample *t*-tests of each resource display condition indicated that, in trials which did not include the resource display, participants rated their workload as lower in the augmented communication condition compared to the standard condition, $t\ (15) = 3.11$, $p < .05$. However, no such difference was found between communication conditions in trials with access to the resource display ($p > .05$). The results of these analyses may indicate that the benefits of access to the virtual whiteboard, in terms of workload reduction, are relatively weak and may be annulled by increased task demands associated with divided attention (as described previously with regards to the auditory monitoring task performance). The relationship between team communication and resource display conditions is illustrated in Figure 10.

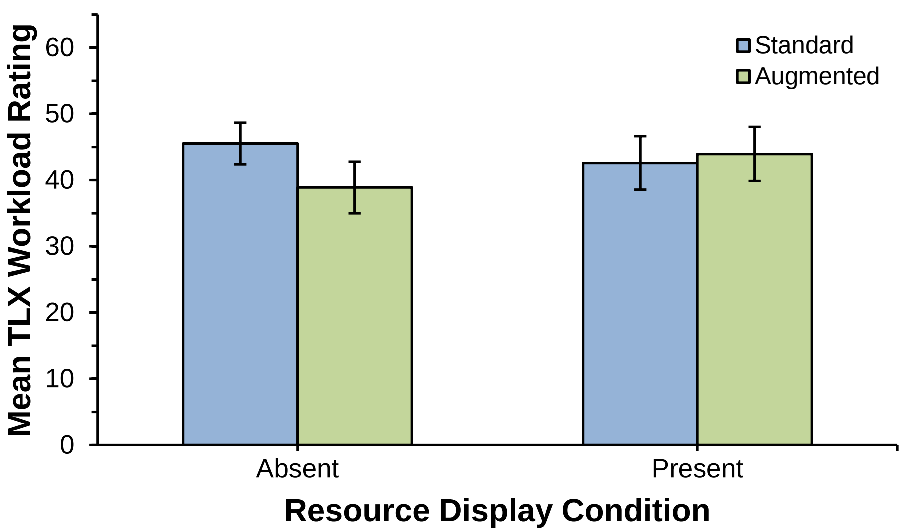

Figure 9. Mean NASA-TLX workload ratings as a function of team communication and resource display conditions. Error bars are standard errors.

2.3.4.2. Modified-TLX Workload

The mean M-TLX workload rating, computed across subscales, reported in this experiment was 31.20 (*SE* = .96). As was the case with the NASA-TLX, this value is near the midpoint of the scale, indicating that participants found the ABM task to be moderately to highly demanding.

ANOVA analysis of the M-TLX workload ratings revealed a statistically significant main effect of *M-TLX subscale*, F (2.17, 32.57) = 3.82, $p < .05$. In addition, statistically significant interactions were found between *team communication* and *resource display* conditions, F (1, 15) = 18.97, $p < .05$, and between *resource display* condition and *M-TLX subscales*, F (3.01, 45.17) = 3.98, $p < .05$. No other sources of variance in the analysis were significant (all $p > .05$).

In exploring the *team communication* × *resource display* interaction, follow-up simple main effects paired-sample *t*-tests for each team communication condition indicated that, in augmented communication trials, participants rated their workload as modestly lower when the resource display was absent compared to when it was present, t (15) = -3.51, $p < .05$. However, no such difference was found between resource display conditions in standard communication trials ($p > .05$). This relationship is illustrated in Figure 11. These results seem to further support previous assertions concerning a reduction of benefits from the virtual whiteboard with divided attention.

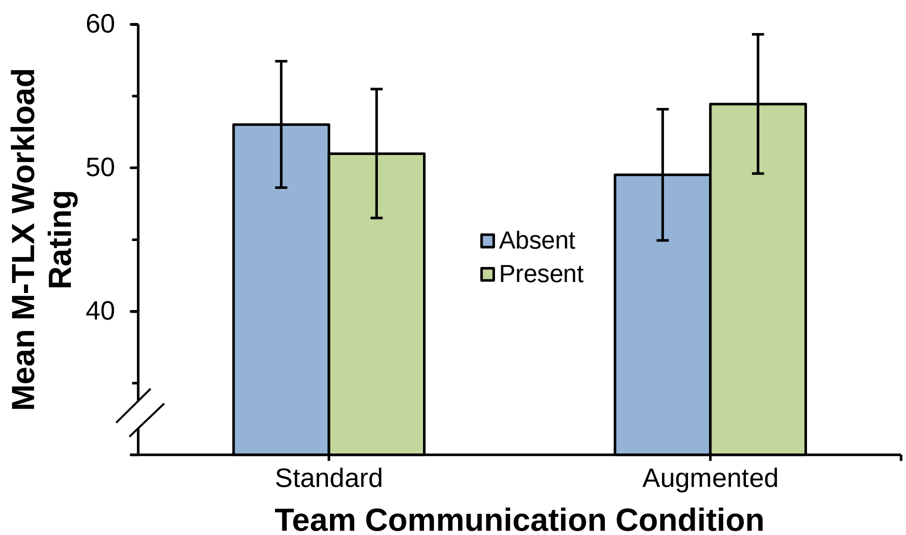

Figure 10. Mean M-TLX workload ratings as a function of team communication and resource display conditions. Error bars are standard errors.

To further explore the *resource display* × *M-TLX subscale* interaction, paired-sample *t*-tests were computed comparing the resource display present and absent conditions for each M-TLX subscale. However, these analyses revealed no statistically significant differences between resource display conditions (all comparisons $p > .05$). Though not statistically significant, examination of Figure 12 suggests that participants tended to rate the communication and leadership demands of the task as higher when the resource display was present.

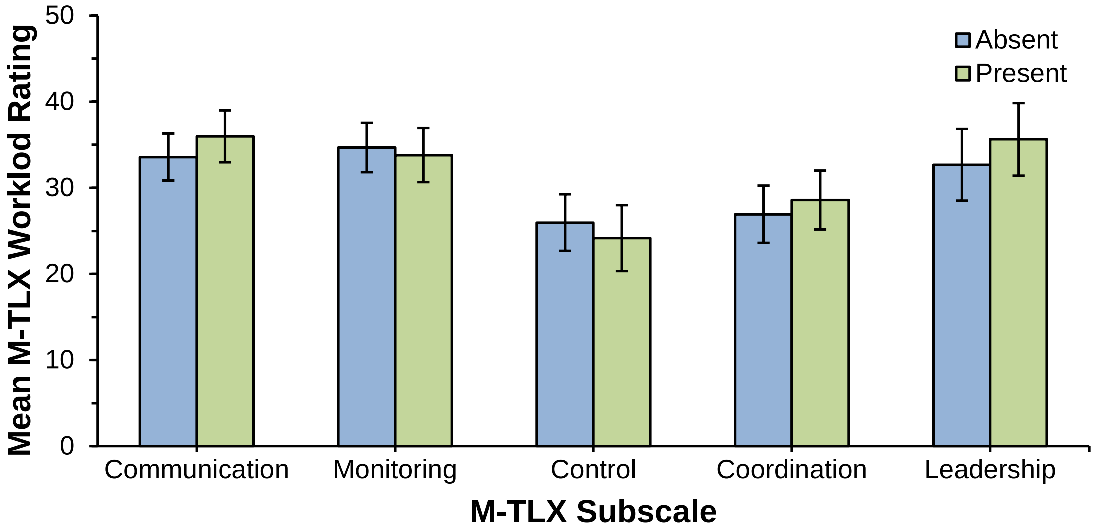

Figure 11. Mean M-TLX workload ratings as a function of resource display condition and M-TLX subscale. Error bars are standard errors.

2.3.4.3. M-MRQ Workload Profile

The mean M-MRQ workload rating, computed across subscales, reported in this experiment was 45.72 (*SE* = 1.64). Again, this value is near the midpoint of the scale, further supporting the assertion that participants found the ABM task to be moderately to highly demanding.

ANOVA Analysis of the M-MRQ workload data indicated a statistically significant main effect of *MRQ subscale*, F (4.82, 72.29) = 17.91, $p < .05$. No other sources of variance in the analysis were significant (all $p > .05$). As is illustrated in Figure 13, participants' M-MRQ ratings appear to be driven by the *auditory linguistic* (AL), *short-term memory* (STM), *spatial attentive* (SA), *visual temporal* (VT), and *vocal process* (V) subscales.

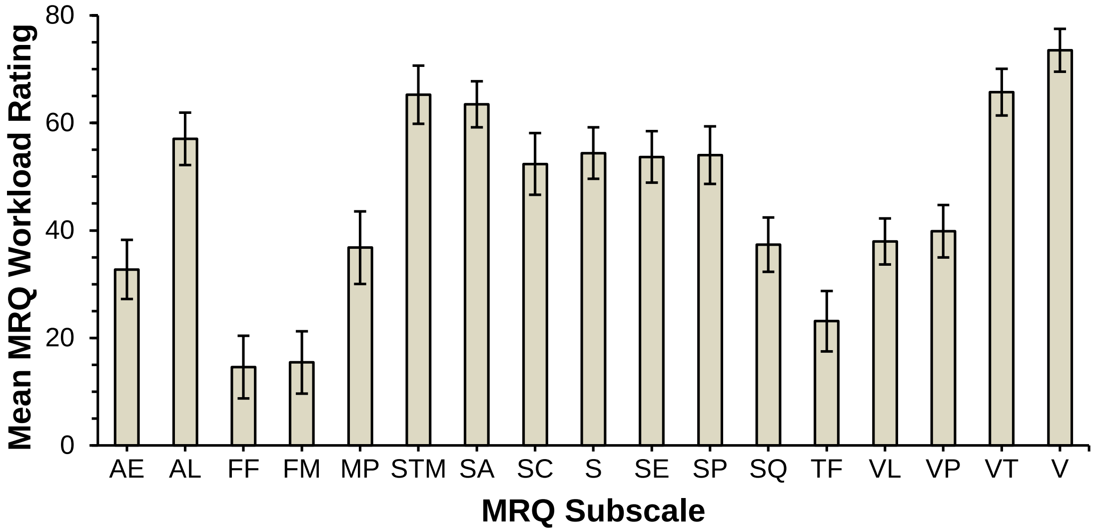

Figure 12. Mean workload ratings for each of the MRQ subscales. Error bars are standard errors.

2.4. Discussion

The purpose of this experiment was to investigate the effects of two collaboration technologies, a virtual whiteboard and a resource display, on team performance, communication, and workload in a simulated command and control task. In pursuit of this goal, novice operators performed C2 tasks in a laboratory-based ABM simulation. It was originally hypothesized that access to these technologies should reduced reliance on oral communication by providing teams with additional communication channels. The availability of the collaboration tools was also expected to improve performance of the air defense and auditory monitoring tasks and plausibly to reduce operator workload. While these hypotheses were generally supported, the results of the experiment were more nuanced than anticipated.

As predicted, access to the virtual whiteboard and resource display significantly reduced the frequency of oral communication. A content analysis of the team communication data indicated that the observed reductions in oral communication were consistent with the types of information each technology was designed to convey. In addition, access to the virtual whiteboard reduced the total duration of team communication. Overall, these results support the view that collaboration technologies may provide an effective, alternative means for team communication. Contrary to initial predictions, the virtual whiteboard and resource display did not significantly alter several indices of team performance in the air defense task, with the exception that teams prosecuted enemy targets more quickly in trials with access to the whiteboard. Performance of the auditory monitoring task, on the other hand, was generally improved by access to the virtual whiteboard, though the observed benefits were reduced by the simultaneous presentation of the resource display. Finally, operator workload was diminished by the virtual whiteboard, but the observed diminution was relatively fragile, in that simultaneous presentation of the resource display resulted in workload levels similar to those reported without the whiteboard.

2.4.1. Virtual Whiteboard

C2 operators are frequently required to communicate using overloaded radio channels within a field of moderate to high ambient noise (Bolia et al., 2005). The results of Experiment 1 tend to support previous research on the utility of a virtual whiteboard for communication in command and control environments (Schwartz et al., 2008; Vincent et al., 2009). Inclusion of the whiteboard yielded modest gains in team performance and more substantial improvement on the auditory monitoring task, but more importantly, reduced the frequency of oral (radio) communication and reduced operator workload. By offloading (or potentially supplementing) some oral communication with whiteboard marks, teams were better able to engage the auditory monitoring task and aspects of the air defense task. This demonstrates that it is possible for personnel to successfully communicate critical information through a non-verbal medium without a concomitant reduction in task performance.

However, the benefits of the virtual whiteboard in this experiment were relatively brittle, in that concurrent presentation of the resource display reversed those gains and returned performance and workload to levels observed on trials without access to the whiteboard. This suggests that participants in this experiment may have experienced some difficulty when dividing attention across displays. The need for participants to monitor the DRAW and tactical displays, which were relatively well integrated (one overlaid the other), and simultaneously monitor and extract information from the resource display may have been sufficiently attentionally demanding to negate the benefits of the virtual whiteboard.

2.4.2. Resource Display

The results of this experiment also support previous research on the efficacy of a resource display as a means to disseminate crucial information without reliance on oral communication (Schwartz et al., 2008). Access to the resource display successfully reduced the frequency of oral communication without adversely affecting team performance or workload. However, the observed interaction between the virtual whiteboard and the resource display suggests that participants may have had difficulty dividing attention across displays or extracting information from the resource display in a timely fashion (or both). These possibilities suggest two solutions:

Firstly, following the recommendations of Wickens and Carswell (1995; see also Flach & Bennett, 1996, for a discussion of these issues), information from the resource display could be integrated into the primary tactical display by presenting weapon and fuel information with asset icons, allowing operators to more rapidly integrate and assimilate spatial location and status information. However, inclusion of this additional information may quickly lead to undesired screen clutter, suggesting that operators may benefit from a control to display or hide the data. Secondly, aspects of the information conveyed by the resource display could be depicted in an analog, rather than digital, format. As noted by Grether (1949) and others (e.g., Tole, Stephens, Harris, & Ephrath, 1982; Wickens & Hollands, 2000), digital presentation of information may lead users to mentally transform that information to an analog conceptual representation, imposing an additional processing step and potentially leading to longer visual fixations, longer processing time, and a greater probability of error. For example, fuel information was represented in the resource display in a “minutes remaining” format, which steadily decreased over time. This style of representation required participants to retain in working memory team assets’ maximum fuel load and minimum time for fueling. A more effective analog alternative could be fuel bars with clearly demarcated maxima and “low fuel” points. Further research exploring these possibilities is clearly warranted.

3.0 EXPERIMENT 2

3.1. Introduction

The purpose of Experiment 2 was to evaluate the impact of the same collaboration tools investigated in Experiment 1 on performance, communication, and workload with ABM domain experts, rather than novice participants. Research with domain experts can provide unique and valuable insight into task performance that is different from, but complimentary to, that of novice participants (e.g., Ericsson & Williams, 2007). Subject matter experts are likely to possess a deeper and more nuanced understanding of C2 operations that may then influence their task strategies and utilization of the collaboration tools investigated.

Given these anticipated differences, some qualitative and quantitative disparities in experimental outcomes were expected between the novice participants in Experiment 1 and the domain experts of Experiment 2 (e.g., with regard to the frequency and content of team communications, subjective workload responses, etc.). Overall, however, comparable results were predicted between the two experiments; similarity in performance, communication, and workload trends in Experiment 2 would help to validate the results of Experiment 1, and support the utility of novel collaboration technologies in C2 environments.

Specifically, based on the results of Experiment 1, it was hypothesized that the availability of the virtual whiteboard and resource display would facilitate team performance on the primary ABM task, though the degree of improvement was expected to be relatively modest, and aid performance of the secondary auditory monitoring task. It was also hypothesized that collaboration tool availability would decrease the overall number and duration of radio transmissions, and that reductions in radio communication would be reflected in semantic categories associated with information conveyed by the collaboration tools (e.g., move and attack directives, resource information, etc.). Finally, it was hypothesized that operator workload would be diminished with access to the collaboration tools, but that reductions could be offset by demands associated with time-sharing attention across displays in some conditions.

3.2. Methods

3.2.1. Participants

Nineteen men and one woman between the ages of 26 and 47 years old ($M = 35.45$, $SD = 6.19$) served as participants in this experiment. Participants were drawn from AWACS crews of the 605th AWACS Test and Evaluation squadron. They had an average of 14.98 years of military experience ($SD = 6.68$ years), and an average of 6.93 years of AWACS experience ($SD = 6.03$ years). All participants volunteered for the study and were not compensated for their participation. In addition, as in Experiment 1, the roles of the sweep and tanker operators were filled by three confederates of the experimenters. Confederates were compensated for their participation. In total, the experimental sample included ten teams; each team consisted of two participants and three confederates.

3.2.2. Experimental Design

The participants employed in this experiment had relatively limited availability (approximately one hour) due to the constraints of their normal military duties, which necessitated some alteration of the experimental design for Experiment 2. As such, a mixed design was adopted, featuring two resource display conditions (absent, present) combined factorially with two communication conditions (standard, augmented) and two auditory monitoring task conditions (absent, present). Resource display condition was a between-subjects factor, and communication and auditory monitoring task conditions were within-subjects factors. Each experimental team completed four mission trials. Team communication condition was a blocked factor. Within each block, half of the participant teams first completed one trial in the *auditory monitoring task absent* condition, followed by a trial in the *monitoring task present* condition; the remaining participant teams experienced those conditions in reverse order. The presentation order of the team communication and auditory monitoring task factors was counterbalanced across teams.

Dependent measures included in Experiment 2 comprised indices of team performance in the simulated air defense task, performance on the auditory monitoring task, frequency and content of team communications, and a single measure of subjective workload (the NASA-TLX).

3.2.3. Apparatus

3.2.3.1. Workstations

Ten notebook computers were required in this experiment. A complete list of the hardware specifications for each computer is displayed in Table 6. Of these computers, five Toshiba tablet PCs were employed as workstations for the participants and confederates, and were outfitted with a standard mouse and a secondary Dell 1703FPs 17 inch LCD monitor. For participant WDs, the Toshiba's 12.1 inch monitor was disabled and the Dell monitor displayed the DDD and DRAW software applications. Conversely, the confederates' workstations were configured so that DRAW was presented on the Toshiba monitor and the DDD software was presented on the Dell monitor. The sixth Toshiba tablet PC was employed as an "observer" station for the experimenters and also hosted software which allowed the experimenters to implement the conditions of each trial. The Sony Vaio notebook acted as the experiments' domain controller, and as the DHCP and SQL servers. The three Gateway notebook PCs were used to play a recording of pink noise (produced through a pair of Optimus Pro 77 speakers), and hosted additional data recording software. All of the computers employed in this experiment were networked using a Netgear GS748T gigabit switch which provided standard TCP/IP Ethernet connectivity.

Table 6. Hardware specifications for the ten computers employed in Experiment 2.

Computer	Quantity	Manufacturer	Model	Processer	Operating System	RAM	Network
Participant and observer workstations	6	Toshiba	Protégé M200	Intel Pentium M 1.7 GHz	MS XP Professional	2 GB	1 Gbps
Server and data recording	3	Gateway	M675PRR	Intel Pentium 4 HT 3.2 GHz	MS XP Professional	1 GB	1Gbps
Domain Controller, DHCP and SQL Server	1	Sony	Vaio PCG-GRT390Z	Intel Pentium 4 HT 3.2 GHz	MS Windows Server 2003	2 GB	1 Gbps

Note. MS = Microsoft.

During the experiment, teammates communicated with each other using Sennheiser Binaural PC headsets (model PC155). These headsets feature noise canceling .8 inch microphones and were directly connected to the microphone and headphone outputs of the Toshiba tablet PCs.

3.2.3.2. Synthetic Task Environment

Experiment 2 utilized the same DDD air defense task employed in Experiment 1. All task parameters were identical with the exception of trial duration; practice and experimental trials were limited to seven minutes in Experiment 2 to accommodate operators' scheduling constraints and limited availability.

3.2.3.3. Supporting Software

The supporting software (e.g., DRAW, the resource display, the auditory monitoring task, etc.) employed in Experiment 2 was identical to that of Experiment 1.

3.2.4. Questionnaires

Due to the time constraints of the participants, WDs completed only a single measure of subjective workload, the NASA-TLX (Hart & Staveland, 1988), following each trial. As in Experiment 1, confederates were not required to complete the questionnaire.

3.2.5. Procedure

As in Experiment 1, the sweep and tanker operators were experimental confederates. These confederates completed the same behavioral and task training that confederates in Experiment 1 did.

Prior to experimental data collection, participant WDs completed a short, 15-minute training session. During this time, they received instruction on the DDD simulation, the radio software, DRAW, and the resource display. Participants did not require brevity training, in that their normal duties provided more than sufficient preparation.

Participants were informed that the purpose of the study was to evaluate how anticipated communication technologies may impact operator and team performance, and that they would be engaged in a medium-fidelity AWACS simulation which required teamwork to meet the scenario's objectives. They were further instructed that the performance of the team would be scored following each trial for how well they had met their objectives and followed the rules of the simulation (as described in Experiment 1). Teams then completed one practice trial, which allowed them to further familiarize themselves with the task and collaboration tools employed in the experiment.

Next, teams were assigned an order of presentation of the experimentally manipulated factors. The experimental schedule of conditions was counterbalanced across teams to control order effects. During the experimental data collection, teams completed four trials, one in each experimental condition. Data collection was completed in approximately one hour. During trials which included the auditory monitoring task, messages were broadcast to the WDs every 30 seconds (different target and distracter messages were sent to each participant), for a total of 14 target messages per trial. As in Experiment 1, DDD simulation events and Morae recordings were logged for later analysis.

3.3. Results

3.3.1. Team Performance

As in Experiment 1, the DDD software recorded the team score, the number of enemy aircraft intercepted, the total time required to prosecute an enemy aircraft, the percentage of enemy aircraft that successfully penetrated friendly airspace, and the number of team assets lost during each experimental trial. To examine the effects of the experimental manipulations on team performance in Experiment 2, the mean was calculated for each team on each variable. These values, presented in Table 7, were then tested for statistically significant differences using separate 2 (resource display) × 2 (team communication) × 2 (auditory monitoring task) mixed model ANOVAs.

The results of these analyses indicated that teams in the *resource display present* condition were significantly more successful at the air defense task across several indices of performance than teams in the *absent* condition. Teams with access to the *resource display* achieved higher overall team scores, $F(1, 8) = 7.79$, $p < .05$, intercepted more enemy aircraft, $F(1, 8) = 7.18$, $p < .05$, had shorter prosecution times, $F(1, 8) = 5.99$, $p < .05$, and allowed fewer aircraft to penetrate friendly airspace, $F(1, 8) = 7.23$, $p < .05$. No other sources of variance in the analyses were significant (all $p > .05$). The observed benefit of the resource display in this experiment may be due to reduced cognitive load associated with asset management and planning, which allowed operators to focus more fully on other aspects of task performance.

Table 7. Mean team performance across several task indices as a function of resource display, team communication and auditory monitoring task conditions.

		Performance Variables			
Trial Condition	Team Score	Enemy Aircraft Intercepted	Time to Prosecute	Airspace Penetration	Team Assets Lost
RD Absent					
Standard					
AMT Absent	68.92 (3.62)	16.80 (.97)	131.49 (6.44)	45.92 (6.44)	3.20 (.37)
AMT Present	63.04 (5.57)	13.60 (.98)	141.53 (8.83)	41.55 (3.87)	3.40 (1.03)
Augmented					
AMT Absent	73.77 (4.53)	16.00 (.77)	133.29 (5.13)	40.95 (4.81)	2.40 (.93)
AMT Present	61.06 (7.95)	14.20 (1.59)	137.25 (10.37)	33.84 (7.00)	4.00 (.89)
RD Present					
Standard					
AMT Absent	79.73 (4.16)	17.60 (.75)	117.82 (5.52)	24.66 (6.59)	1.80 (.86)
AMT Present	72.87 (6.15)	16.80 (1.24)	116.68 (4.20)	29.62 (4.94)	2.60 (.68)
Augmented					
AMT Absent	80.38 (4.45)	18.00 (.95)	117.81 (10.91)	21.12 (7.92)	1.40 (.68)
AMT Present	80.43 (5.89)	18.20 (1.39)	118.47 (11.14)	26.79 (7.18)	1.40 (.60)

Note. RD = Resource display. AMT = Auditory monitoring task. Values in parentheses are standard errors.

3.3.2. Auditory Monitoring Task Performance

As in Experiment 1, the CRM program recorded the number of signals responded to and the number of correct responses each participant made in each trial. The mean number of overall and correct responses were analyzed for statistically significant difference between conditions using separate 2 (resource display) × 2 (team communication) mixed model ANOVAs.

The response rate to the auditory monitoring task in this experiment was relatively low (participants responded to approximately 39% of the signals). Across conditions, the mean number of operator responses per trial was 5.43 (SE = .78), and the mean number of correct responses was 2.90 (SE = .42). The results of the ANOVA analyses revealed no statistically significant differences between conditions on either auditory task performance variable (all main effects and interactions $p > .05$).

3.3.3. Team Communication

Following the completion of experimental data collection, audio recordings and DRAW logs of the communications between teammates were compiled and examined. Across trials, teams sent an average of 120.73 radio messages per trial. In addition, when the virtual whiteboard was available, teams sent an average of 57.70 DRAW messages per trial. As a manipulation check, the mean numbers of DRAW marks sent per trial were tested against a value of zero using a one-sample *t*-test to establish that teams were, in fact, using the tool. The results of this analysis indicated that participants were communicating at a rate greater than zero using DRAW marks, t (9) = 20.98, $p < .05$.

3.3.4. Virtual Whiteboard Communication

To examine the number of virtual whiteboard communications sent for potential differences related to the experimental manipulations, the mean numbers of DRAW messages sent in each augmented communication trial were computed and compared using a 2 (resource display) × 2 (team communication) mixed model ANOVA. The results of the analysis indicated that teams sent approximately the same number of DRAW communications in each condition (all main effects and interactions $p > .05$).

3.3.4.1. Radio Communication

The frequencies and durations of team communication during each trial were computed as described in Experiment 1. Mean values were calculated for each team and experimental condition, and these values were tested for statistically significant differences between conditions using separate 2 (resource display) × 2 (team communication) × 2 (auditory monitoring task) mixed model ANOVAs. For the frequency of radio communication, statistically significant main effects were detected for the *team communication*, F (1, 8) = 147.32, $p < .05$, and *auditory monitoring task* conditions, F (1, 8) = 6.78, $p < .05$. No other sources of variance in the analysis were significant (all $p > .05$). As is depicted in Figure 14, participants made significantly fewer radio communications during trials with access to the virtual whiteboard and trials featuring the auditory monitoring task.

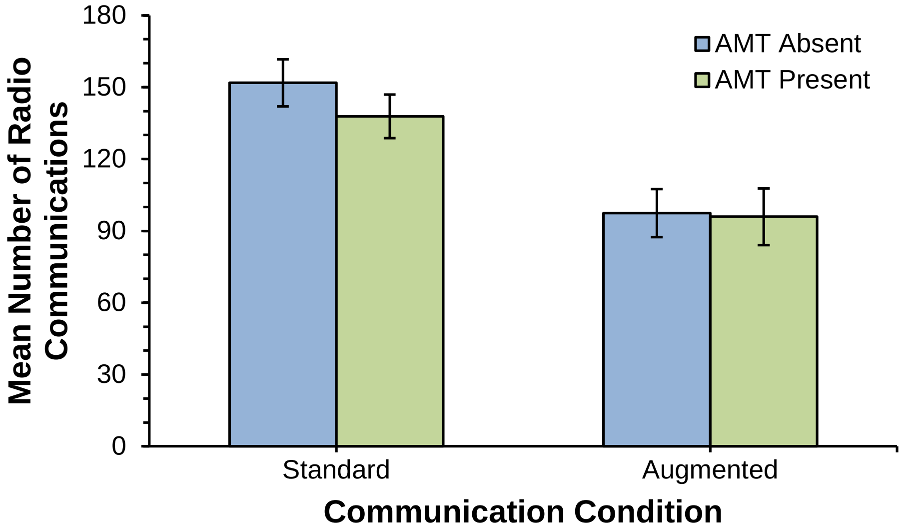

Figure 13. Mean number of radio communications as a function of team communication and auditory monitoring task conditions. Error bars are standard errors.

For the duration of radio communications, a statistically significant main effect of *team communication* condition was detected, F (1, 8) = 94.25, p < .05. No other sources of variance in the analysis were significant (p > .05). The average duration of all radio communication during a trial was approximately 70% longer in the standard communication condition (M = 330.13 s, SE = 23.83 s) compared to the augmented condition (M = 196.81 s, SE = 23.22 s). Overall, the observed reductions in frequency and duration of radio communication in the augmented condition suggest that teams were transitioning communication from the radio channel to the virtual whiteboard. In addition, the decrement in communication associated with the auditory monitoring task suggests that operators were attempting to engage that task.

3.3.4.2. Radio Communication Content Analysis

A separate sampling strategy was selected for the content analysis in Experiment 2. In this experiment, rather than transcribing and coding all radio communications, a random sample of 1,000 communications was selected for inclusion in the analysis. The subset was constructed such that approximately 30 statements were sampled from each of 33 trials (this is analogous to sampling approximately 25% of communications from 83% of all trials), and included 200 statements sampled from each team position.

The categorization scheme and coding process employed was the same as in Experiment 1. The interrater reliability of the two judges, assessed by the proportion of overall agreement and Cohen's kappa, was again deemed by the authors to be sufficient (proportion of overall agreement = .83; Cohen's kappa = .78, p < .05). The percentage of radio communications in each category for each experimental condition is presented in Table 8. As can be observed in the table, access to the resource display resulted in relatively substantial decreases in the percentage of radio communications classified as *resource status – update* and *resource status – request*, which is consistent with the information conveyed by the display, and increases in the number of *clarification / confirmation* and *situation update* communications. Access to the virtual whiteboard in the augmented communication condition resulted in decrements in the percentage of radio communications classified as *directive – attack* and *directive – resupply*, which is consistent with the types of messages the DRAW was designed to convey, and an increase in *clarification / confirmation* communications. Finally, addition of the auditory monitoring task did not appear to strongly influence the content of participants' communications in this sample.

Table 8. Percentage of radio communications by category as a function of team communication and resource display conditions.

	RD Absent				RD Present				
	Standard		Augmented		Standard		Augmented		
Category[a]	AMT Absent	AMT Present	AMT Absent	AMT Present	AMT Absent	AMT Present	AMT Absent	AMT Present	% of Total[b]
Clarification / Confirmation	20.83	29.30	27.92	27.91	21.82	25.96	46.67	48.48	31.08
Situation Update	20.14	15.29	19.29	19.38	29.09	25.00	31.11	32.32	22.75
Resource Status – Update	24.31	19.75	19.29	21.71	9.09	17.31	15.56	11.11	18.33
Resource Status – Request	15.97	15.29	16.24	16.28	3.64	1.92	.74	.00	10.29
Directive – Resupply	12.50	9.55	4.57	6.98	21.82	11.54	.00	1.01	7.45
Directive – Attack	5.56	8.28	5.58	3.10	12.73	14.42	.00	.00	5.69
Coordinate	.69	.64	4.57	2.33	1.82	2.88	5.93	6.06	3.14
Social	.00	1.91	2.54	2.33	.00	.96	.00	1.01	1.27

Note. RD = Resource display. AMT = Auditory monitoring task.

[a]Categories are presented in their order of predominance, from largest to smallest, in the 1,000 item sample set.

[b]Indicates the prevalence of communications in each category from the sample set, collapsed across experimental conditions to facilitate cross-condition comparisons.

Of some interest is the difference in categorical predominance observed between the communications of the novices in Experiment 1 (Table 5) and the domain experts (Table 8). Domain experts demonstrated a great deal more concern about the state of the simulation (*situation update*) and of their assets (*resource status – update* and *resource status – request*), as indexed by the larger percentages of total communications in each of those categories. By contrast, novices were more focused on maneuvering assets and attacking (*directive – attack*). These differences may indicate a fundamental divergence in approaches to the air defense task, as domain experts employed a strategy reliant on maintaining situation awareness and novices displayed a more aggressively oriented strategy.

3.3.5. NASA-TLX Workload

As discussed previously, due to the limited availability of the participants in this experiment the M-TLX and MRQ were omitted, leaving the NASA-TLX as the sole measure of workload employed. Additionally, due to a computer error the response data of four participants was lost and could not be recovered. Mean ratings for the six NASA-TLX subscales were computed for the remaining eight participants; the mean TLX workload rating, computed across subscales, reported in this experiment was 57.81 (SE = 1.64). This value is above the midpoint of the scale, indicating that the participants found the ABM task to be highly demanding.

Differences in workload ratings for each condition were tested for statistical significance using a 2 (resource display) × 2 (team communication) × 2 (auditory monitoring task) × 6 (TLX subscale) mixed model ANOVA. The results of the analysis indicated statistically significant main effects of *resource display* condition, F (1, 14) = 5.01, $p < .05$, and *TLX subscale*, F (3.07, 42.96) = 20.08, $p < .05$. Participants in the *resource display present* condition rated their workload as lower (M = 53.60, SE = 2.62) compared to participants in the *absent* condition (M = 64.83, SE = 4.84). Depicted in Figure 15, participants' workload estimates appear to be driven by the mental demand, temporal demand, and effort associated with task performance (as was the case for the novice operators in Experiment 1).

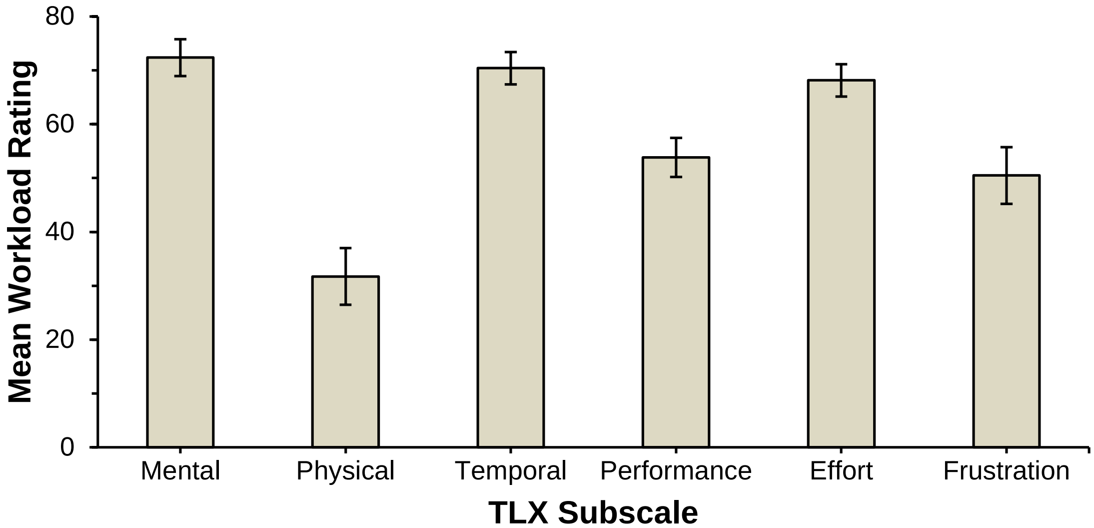

Figure 14. Mean workload ratings for each of the NASA-TLX subscales. Error bars are standard errors.

3.4. Discussion

The purpose of Experiment 2 was to investigate the effects of a virtual whiteboard and a resource display on team performance, communication, and workload in a simulated C2 task with domain experts, rather than novice participants. From the results of Experiment 1, it was hypothesized that the availability of the virtual whiteboard and resource display would facilitate team performance on the primary ABM task, though the degree of improvement was expected to be relatively modest, and aid performance of the secondary auditory monitoring task. It was also hypothesized that collaboration tool availability would decrease the overall number and duration of radio transmissions, and that reductions in radio communication would be reflected in semantic categories associated with information conveyed by the collaboration tools (e.g., move and attack directives, resource information, etc.). Finally, it was hypothesized that operator workload would be diminished with access to the collaboration tools, but that reductions could be offset by demands associated with time-sharing attention across displays in some conditions. These hypotheses were generally supported, though domain experts appeared to derive greater benefit from access to the resource display than did the novice participants of Experiment 1.

As was observed with novices, access to the virtual whiteboard significantly reduced the frequency and duration of oral communication. A content analysis of the communication data indicated that, as in Experiment 1, the observed reduction was consistent with the types of information the whiteboard was designed to convey. Overall, these results again support the view that collaboration technologies may provide an effective, alternative means for team communication.

Contrary to the results observed with novices, across several indices of performance in the air defense task (i.e., team score, number of enemy aircraft killed, time to prosecute enemy aircraft, friendly airspace penetration), domain experts' performance was modestly improved with access to the resource display. However, access to collaboration technologies did not alter their performance on the auditory monitoring task (though the observed decline in communications during trials featuring the auditory monitoring task suggests that participants were indeed attempting to engage that task). Finally, domain experts with access to the resource display rated their workload as lower than those without access.

3.4.1. Virtual Whiteboard

The results of Experiment 2 tend to further support previous research on the utility of a virtual whiteboard for communication in command and control environments (Schwartz et al., 2008; Vincent et al., 2009). As was the case in Experiment 1, inclusion of the whiteboard reduced the frequency and duration of oral (radio) communication. Though differences were observed in the effects of the whiteboard on novice and domain expert participants' performance, both groups evidenced a significant decrement in reliance on oral communication with no concomitant reduction in task performance or increase in subjective workload. As this pattern of results was obtained with both novice and domain expert participants, it further demonstrates the likelihood that C2 personnel could successfully communicate critical, task-relevant information through a non-verbal medium without adversely impacting team effectiveness.

3.4.2. Resource Display

The results of this experiment also support previous research on the efficacy of a resource display for communicating information in distributed team environments (Schwartz et al., 2008). Access to the display reduced the frequency of oral communication and engendered substantive benefits for domain experts, including improved task performance and reduced workload. Without contradicting the design suggestions offered previously (in Section 2.4.2), the observed enhancement of domain experts' performance with access to the resource display (which contrasts with the effects of the display on novice participants' performance) may be due to the task proficiencies of those participants. As noted by Knott et al. (2006), AWACS personnel are routinely required to divide attention across multiple information sources in performance of their duties. This includes simultaneous monitoring of several information channels (e.g., tactical displays, radio channels, chat rooms, etc.) with appropriate responses to each as the need arises. As has been found in other domains such as aviation (e.g., Bellenkes, Wickens, & Kramer, 1997) and driving (e.g., Wikman, Niemeinen, & Summala, 1998), domain experts in this experiment may have been more adept at rapidly transitioning attention between displays and extracting task-critical information from those sources, allowing them to benefit to a greater extent from the information conveyed by the resource display.

4.0 GENERAL DISCUSSION

The results of Experiment 1 and Experiment 2 coincide with previous research supporting the utility of collaboration technologies as alternative modes of team communication in C2 environments (e.g., Schwartz et al., 2008; Vincent et al., 2009). While access to these collaboration technologies yielded relatively modest improvements in team performance, participant utilization of the tools resulted in substantive reductions in radio communication traffic. This is important since a primary impetus for employing collaboration technologies in military settings is to alleviate reliance on congested radio channels (Knott et al., 2006). Overall, these results indicate that supplemental collaboration technologies are likely to benefit military operators by providing additional, and largely parallel, media for team communication, and by enabling small performance advantages in military operations that may accrue exploitable opportunities for enhancing mission success. In addition, the results demonstrate that collaboration technologies may not *necessarily* impose additional workload on operators associated with monitoring those technologies. Still, a thoughtful approach must be taken to ensure that the design and implementation of collaboration technologies in operational settings proceeds in a thoughtful manner; this is likely to require field research to determine appropriate tool format and functionality. Potential avenues for future research include exploring the exact forms that collaboration technologies should take, to whom they should be deployed, and the degree of training and practice operators require to achieve tool proficiency.

5.0 REFERENCES

Alberts, D.S., & Hayes, R.E. (2003). *Power to the edge*. Washington, DC: CCRP Publication Series.

Baddeley, A.D. (1986). *Working memory*. Oxford University Press, London.

Baltes, B.B., Dickson, M.W., Sherman, M.P., Bauer, C.C., & LaGanke, J.S. (2002). Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes, 87*, 156-179.

Bellenkes, A.H., Wickens, C.D., & Kramer, A.F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and menatl model development. *Aviation, Space, and Environmental Medicine, 68*, 569-579.

Boles, D.B., & Adair, L.P. (2001). The multiple resources questionnaire (MRQ). *Proceedings of the Human Factors and Ergonomics Society, 45*, 1790-1794.

Bolia, R.S., Nelson, W.T., Ericson, M.A., & Simpson, B.D. (2000). A speech corpus for multitalker communications research. *Journal of the Accoustical Society of America, 107*, 1065-1066.

Bolia, R.S., Nelson, W.T., Vidulich, M.A., Simpson, B.D., & Brungart, D.S. (2005). Communications research for command and control: Human-machine interface technologies supporting effective air battle management. *Proceedings of the 10th International Command and Control Research and Technology Symposium*. Washington: Command and Control Research Program. Retrieved from http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/279.pdf

Boiney, L. (2005). Team decision making in time-sensitive environments. *Proceedings of the 10th International Command and Control Research and Technology Symposium*. Washington, DC: Command and Control Research Program. Retrieved from http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/175.pdf

Bolstad, C.A. & Endsley, M.R. (2005). Choosing team collaboration tools: Lessons from disaster recovery efforts. *Ergonomics in Design, 13*, 7-14.

Bowers, C., Braun, C., & Morgan, B. (1997). Team workload: Its meaning and measurement. In M.T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 85-108). Mahwah, NJ: Erlbaum.

Cohen, J. (1960). A coefficient for agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cooke, N.J., & Shope, S.M. (2004). Designing a synthetic task environment. In S.G. Schiflett, L.R. Elliott, E. Salas, & M.D. Coovert (Eds.), *Scaled worlds: Development, validation and applications* (pp. 263-278). Brookfield, VT: Ashgate.

DDD_Console [Computer software]. (2006). Dayton, OH: Air Force Research Laboratory.

DDD_Results [Computer software]. (2006). Dayton, OH: Air Force Research Laboratory.

DISlog (Version 14) [Computer software]. (1996). Huntsville, AL: CG^2, Inc.

Dynamic real-time animated whiteboard (DRAW) [Computer software]. (2006). Dayton, OH: Air Force Research Laboratory.

Ericsson, K. A., & Williams, A.M. (2007). Capturing naturally occurring superior performance in the laboratory: Translational research on expert performance. *Journal of Experimental Psychology: Applied*, *13*, 115-123.

Finomore, V.S., Knott, B.A., Nelson, W.T., Galster, S.M., & Bolia, R.S. (2007). The effects of multimodal collaboration technology on subjective workload profiles of tactical air battle management teams. *Proceedings of the International Symposium on Aviation Psychology, 14*, 190-196.

Finomore, V.S., Warm, J.S., Matthews, G., Riley, M.A., Dember, W.N., Shaw, T.H., Ungar, N.R., & Scerbo., M.W. (2006). Measuring the workload of sustained attention. *Proceedings of the Human Factors and Ergonomics Society, 50,* 1614-1618.

Flach, J.M. & Bennett, K.B. (1996). A theoretical framework for representative design. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and application.* (pp. 65 - 87). Mahwah, NJ: Erlbaum.

Funke, G.J., & Galster, S.M. (2009). The effects of cognitive processing load and collaboration technology on team performance in a simulated command and control environment. *International Journal of Industrial Ergonomics*, *39*, 541-547.

Grether, W.F. (1949). Instrument reading I: The design of long-scale indicators for speed and accuracy of quantitative readings. *Journal of Applied Psychology*, *33,* 363-372.

Hart, S.G., & Staveland, L.E. (1988). Development of a multi-dimensional workload scale: Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

Hayes, R.E. (2004). Network centric operations today: Between the promise and the practice. *RUSI Defence Systems*, *7*, 82-85.

Hertel, G., Geister, S., Konradt, U. (2005). Managing virtual teams: A review of current empirical research. *Human Resource Management Review, 15*, 69-95.

Jeziorski, A. (2008, October 01). Boeing shows off more potential AWACS mods. *C4ISR Journal*. Retrieved from http://www.c4isrjournal.com/story.php?F=3665858

Kirk, R.E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Knott, B.A., Bolia, R.S., Nelson, W.T., Galster, S.M. (2006). Effects of collaboration technology on the performance of tactical air battle management teams. *Proceedings of the Symposium on Human Factors Issues in Network-Centric Warfare*, Sydney, Australia.

MacMillan, J., Entin, E.B., Hess, K.P., & Paley, M.J. (2004). Measuring performance in a scaled world: Lessons learned from the Distributed Dynamic Decisionmaking (DDD) synthetic team task. In S.G. Schiflett, L.R. Elliott, E. Salas, & M.D. Coovert (Eds.), *Scaled worlds: Development, validation and applications* (pp. 154-180). Brookfield, VT: Ashgate.

Microsoft Virtual PC 2004 [Computer software]. Redmond, WA: Microsoft Corporation.

ModIOS (Version 2.3.4) [Computer software]. (2002). Scottsdale, AZ: General Dynamics C4 Systems.

Morae (Version 1.3) [Computer software]. (2005). Okemos, MI: Techsmith Corporation.

Muller, K.E., & Barton, C.N. (1989). Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association, 84,* 549–555.

Parasuraman R., & Riley, V. (1997). Humans and automation: Use, misuse, abuse. *Human Factors*, *39*, 230-253.

Pharmer, J., Cropper, K., McKneely, J., & Williams, E. (2004). *Tactical Tomahawk weapon control system V6 land attack combat system prototype human-computer interface: Test report for FY 03 fleet operability test* (Tech. Rep. No. 3184). San Diego, CA: Naval Air Systems Command, Space and Naval Warfare Systems Center.

PsTools Suite (Version 1.6) [Computer software]. (2006). Redmond, WA: Microsoft Corporation.

Python (Version 2.2.3) [Computer software]. (2003). Hampton, NH: Python Software Foundation.

Red Hat Linux (Version 9.0) [Computer software]. (2004). Raleigh, NC: Red Hat, Inc.

Resource Display [Computer software]. (2006). Dayton, OH: Air Force Research Laboratory.

Salas, E., Cooke, N.J., & Rosen, M.A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors, 50*, 540-547.

Schoeneman, C. (2002). Synergy (Version 1.2.7) [Computer software]. Boston, MA.

Schwartz, D., Knott, B.A., & Galster, S.M. (2008). Effects of a visual communication tool and separable status display on team performance and subjective workload in air battle management. *Proceedings of the 13th International Command and Control Research and Technology Symposium*. Washington: Command and Control Research Program. Retrieved from http://www.dodccrp.org/events/13th_iccrts_2008/CD/html/papers/064.pdf

ScreenCapture [Computer software]. (2003). Dayton, OH: Air Force Research Laboratory.

Sjölander, K., & Beskow, J. (2005). WaveSurfer (Version 1.8.5) [Computer software]. Stockholm, Sweden.

Sloan, D. (2009, July 23). *Boeing awarded production contract for US Air Force AWACS Block 40/45 upgrade*. Retrieved from the Boeing Company web site: http://boeing.mediaroom.com/index.php?s=43&item=761

Subject Survey System (SSS) [Computer software]. (2003). Dayton, OH: Air Force Research Laboratory.

Tatham, S. (2007). PuTTY (Version 0.59) [Computer software]. Cambridge, MA.

Tole, J.R., Stephens, A.T., Harris, R.L., & Ephrath, A.R. (1982). Visual scanning behavior and metnal workload in aircraft pilots. *Aviation, Space, and Environmental Medicine, 53*, 54-61.

Uebersax, J. (2000). *Statistical methods for rater agreement*. Retrieved from http://www.john-uebersax.com/stat/agree.htm

U.S. Department of Defense (2001). *Department of Defense Dictionary of Military and Associated Terms* (Joint Publication 1-02). Washington, DC.

Vidulich, M.A., Bolia, R.S., & Nelson, W.T. (2004). Technology, organization, and collaborative situation awareness in air battle management: Historical and theoretical perspectives. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 233-253). Aldershot, UK: Ashgate Publishing.

Vincent, C., Donnelly, B., Taylor, R., Knott, B., Cottrell, R., Dixon, D., Miller, B., Dukes, A., Bennett, A., & Funke, G. (2009, April). *Adaptive decision support for dynamic mission replanning*. Poster presented at the 15th International Symposium on Aviation Psychology, Dayton, OH.

Warfighter Communication Assessment System (WCAS) [Computer software]. (2005). Dayton, OH: Air Force Research Laboratory.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII*. (pp.239-257). Hillsdale, NJ: Erlbaum.

Wickens, C.D., & Carswell, C.M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, *37*, 473-494.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering psychology and human performance* (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

Wickens, C.D., Vidulich, M., Sandry-Garza, D. (1984). Principles of S-C-R compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. *Human Factors, 26*, 533-543.

Wikman, A., Niemeinen, T., & Summala, H. (1998). Driving experience and time-sharing during in-car tasks on roads of different widths. *Ergonomics, 41*, 358-372.